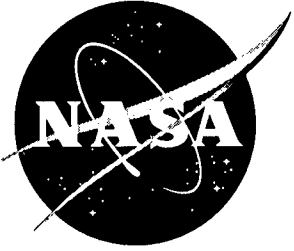


Statistical Technique for Intermediate and Long-Range Estimation of 13-Month Smoothed Solar Flux and Geomagnetic Index

K.O. Niehuss, H.C. Euler, Jr., and W.W. Vaughan



Statistical Technique for Intermediate and Long-Range Estimation of 13-Month Smoothed Solar Flux and Geomagnetic Index

K.O. Niehuss and H.C. Euler, Jr.
Marshall Space Flight Center • MSFC, Alabama

W.W. Vaughan
University of Alabama in Huntsville • Huntsville, Alabama

ACKNOWLEDGMENTS

The authors thank Leonard W. Howell, Jr., Electromagnetics and Aerospace Environments Branch, Robert L. Holland, Computer Sciences Corporation, for technical contributions, Steve Smith, Computer Sciences Corporation, for computer programming support, and Margaret B. Alexander, Electromagnetics and Aerospace Environments Branch, for reviewing this memorandum.

FOREWORD

This document provides the methodology used in the development of the MSFC 13-month smoothed solar flux and geomagnetic index intermediate (months) and long-range (years) statistical estimation technique. These estimates are provided as input to orbital lifetime models.

Section 1 is an introduction and reason the estimates are determined. Section 2 is the history of the development of the computer program that calculates the estimates. Section 3 contains the methodology of the calculation technique. Section 4 describes the flow of the computer program to generate the estimates. Section 5 discusses the engineering use of the computer program's products. Appendix A gives the linear least squares model development. Appendix B is an overview of the cubic connection between the predicted cycle and the mean. Appendix C provides the rationale for the statistics used to calculate the percentile values. Appendix D provides the McNish-Lincoln and quantile computer programs. Appendix E contains an example of the computer program product provided in the monthly Marshall Space Flight Center solar activity memorandum. Appendix F gives an assessment of the estimation technique accomplished for a number of past solar cycles.

Questions or comments, contact NASA Marshall Space Flight Center Systems Analysis and Integration Laboratory, Electromagnetics and Aerospace Environments Branch, Chief, Steven D. Pearson (205) 544-2350.

TABLE OF CONTENTS

1.0	Introduction	1
2.0	Background	1
3.0	MSFC Lagrangian Linear Regression Methodology	2
3.1	Development of $\bar{F}_{10.7}$ and \bar{A}_p Data Bases	2
3.1.1	13-Month Smoothed Solar Flux ($\bar{F}_{10.7}$) Data Base	3
3.1.2	13-Month Smoothed Geomagnetic Index (\bar{A}_p) Data Base	6
3.2	Modified McNish-Lincoln Linear Regression Method	7
4.0	Application of MSFC Lagrangian Linear Regression Technique	8
4.1	$\bar{F}_{10.7}$ and \bar{A}_p Preprocessor Computer Program	10
4.2	Modified McNish-Lincoln Linear Regression Computer Program	12
4.3	Statistical Estimation Output Report Computer Program	14
5.0	Engineering Use of MSFC Solar Activity Statistical Estimation Technique Products	15
6.0	Concluding Remarks	16
7.0	References	17
8.0	Bibliography	18
Appendix A	Linear Least Squares Model Development	A-1
Appendix B	Cubic Connection From Present Predicted Cycle to Mean of Next Cycle	B-1
Appendix C	Estimates of Percentiles for the 13-Month Smoothed Solar Flux Data	C-1
Appendix D	Modified McNish-Lincoln and Quantile Subroutine Examples	D-1
Appendix E	MSFC Monthly Solar Activity Memorandum Example	E-1
Appendix F	MSFC Lagrangian Linear Regression Technique Historical Plots	F-1

LIST OF FIGURES

<u>Figure</u>	<u>Title</u>	<u>Page</u>
2-1	Modified McNish-Lincoln Linear Regression Method.....	2
3-1	\bar{R} versus $\bar{F}_{10.7}$ with Curve Fit from Conversion Equation	4
3-2	13-Month Smoothed (Converted and Observed) Solar Flux ($\bar{F}_{10.7}$)	4
3-3	\bar{C}_i versus \bar{A}_p with Curve Fit from Conversion Equation.....	6
3-4	13-Month Smoothed (Converted and Recorded) Geomagnetic Index (\bar{A}_p).....	7
4-1	MSFC Lagrangian Linear Regression Technique Block Diagram.....	9
4-2	$\bar{F}_{10.7}$ and \bar{A}_p Preprocessor Computer Program Block Diagram.....	11
4-3	Modified McNish-Lincoln Linear Regression Computer Program Block Diagram	13
4-4	Report Program Block Diagram	14
5-1	MSFC Spacecraft Orbital Lifetime Prediction Model Block Diagram.....	15
C-1	Arbitrary Probability Density Function $f(x)$ and Its Cumulative Distribution $F(x)$	C-2
C-2	13-Month Smoothed Lagrangian Flux	C-3
C-3	Mean and Standard Deviation Trend Lines.....	C-3
C-4	Relative Frequency Histogram.....	C-5
C-5	Frequency Polygon	C-5
C-6	Comparison of Frequency Polygon Methods for $Q(0.95)$	C-7
C-7	Comparison of Frequency Polygon Methods for $Q(0.976)$	C-7
C-8	Quantile Plot	C-9
C-9	Interpolated Quantiles	C-10
C-10	Quantile Method	C-11
C-11	Relationship Between the Empirical Distribution Function, F_n , and the Quantile Function, Q	C-12
C-12	Relative Frequency Histogram and Probability Density Function of p_1	C-17
C-13	Relative Frequency Histogram and Probability Density Function for p_{11}	C-17
C-14	Relative Frequency Histogram and Probability Density Function for p_{21}	C-18
C-15	Quantile Method: $Q(0.95)$	C-19
E-1	Plot of Recent Monthly Mean and 13-Month Smoothed Solar Flux.	E-9
E-2	Estimate of 13-Month Smoothed Solar Flux for Balance of Cycle 22, 23, and Beginning of Cycle 24.	E-9
F-1	Minimum to Minimum Cycle 20 MLLRT Model Results Evaluation	F-2
F-2	Maximum to Maximum Cycle 20 MLLRT Model Results Evaluation	F-3
F-3	Minimum to Minimum Cycle 21 MLLRT Model Results Evaluation	F-4
F-4	Maximum to Maximum Cycle 21 MLLRT Model Results Evaluation	F-5
F-5	Minimum to Minimum Cycle 22 MLLRT Model Results Evaluation	F-6
F-6	Maximum to Maximum Cycle 22 MLLRT Model Results Evaluation	F-7

LIST OF TABLES

<u>Table</u>	<u>Title</u>	<u>Page</u>
3-1	Solar Cycle Elements Information in Data Base	5
C-1	Simulated Data and Sample Means and Variances.....	C-16
C-2	Means and Variances of the Beta Distribution	C-16
E-1	Recent Monthly Mean Solar Activity Values.....	E-3
E-2	13-Month Smoothed 10.7-cm Solar Flux	E-4
E-3	Estimates of 13-Month Smoothed Solar Activity for Balance of Cycle 22, Cycle 23, and Beginning of Cycle 24	E-5

LIST OF SYMBOLS AND ACRONYMS

a_p	3-Hourly Value of Geomagnetic Index
A_p	Daily or Monthly Mean Value (depending on use) of Geomagnetic Index
\overline{A}_p	13-Month Smoothed Geomagnetic Index
C	Regression Coefficient
C_i	Magnetic Character Figure
\overline{C}_i	13-Month Smoothed Magnetic Character Figure
CREAT_BLOCK	Create File Subroutine
COMB_YLGINT	Lagrangian Interpolation Subroutine
COV	Covariance
CUBFIT	Cubic Curve Fit Subroutine
EDF	Empirical Distribution Function
$F_{10.7}$	Daily or Monthly Mean Value (depending on use) of 10.7-cm Solar Radio Noise Flux
$\overline{F}_{10.7}$	13-Month Smoothed 10.7-cm Solar Radio Noise Flux
FNDPRCNT	Find the Percentile Subroutine
FORTTRAN	Formula Translation (computer language)
FORECAST	Modified McNish-Lincoln Linear Regression Program
FRMEMO	Memorandum Printout Subroutine
GJR	Inverse Matrix Subroutine
INFLBS	Find-Cubic-Inflection-Point Subroutine
INTERP	Linear Interpolation Subroutine
LINCMC	Modified McNish-Lincoln Subroutine
m	Number of Points per Solar Cycle
MEMO	Memorandum
MERGE	Merge Subroutine
MLLRT	MSFC Lagrangian Linear Regression Technique
MSFC	Marshall Space Fight Center
NASA	National Aeronautics and Space Administration
P	Period Related to Maximum to Maximum or Minimum to Minimum of Solar Cycles
P_j	Period Related to Maximum to Maximum or Minimum to Minimum of j^{th} Solar Cycle
\overline{P}	Mean Period
pdf	Probability Density Function
PCHMAX	Smoothing Subroutine for Maximum to Maximum
PCHMIN	Smoothing Subroutine for Minimum to Minimum

PRCNTILE	Percentiles Subroutine
PREPROC	$\bar{F}_{10.7}$ and \bar{A}_p Preprocessor Program
Q	Quantile
QUANTILE	Quantile Subroutine
R	Wolf's Relative Sunspot Number
\bar{R}	13-Month Smoothed Sunspot Number
\bar{R}_p	Predicted 13-Month Smoothed Sunspot Number
\bar{R}_{mean_p}	Mean of observed 13-Month Smoothed Sunspot Number for n Cycles at Predicted Time Point for \bar{R}_p
$\Delta\bar{R}_p$	Difference between \bar{R}_{mean_p} and \bar{R}_p as Calculated by Modified McNish- Lincoln Linear Regression Method
\bar{R}_o	Observed 13-Month Smoothed Sunspot Number at Prediction Time Point
\bar{R}_{mean_o}	Mean of observed 13-Month Smoothed Sunspot Number for n Cycles at Prediction Time Point for \bar{R}_o
$\Delta\bar{R}_o$	Difference between \bar{R}_{mean_o} and \bar{R}_o at Prediction Time Point
RANKIT	Arrange Data in Ascending Order Subroutine
REPORT	Report Program
t	Time
t_o	Time from Method Initialization Time Point (Prediction Time Point)
VAR	Variance
τ	Mean Time Increment
σ	Standard Deviation

Technical Memorandum

Statistical Technique for Intermediate and Long-Range Estimation of 13-Month Smoothed Solar Flux and Geomagnetic Index

1.0 Introduction

Because no generally accepted solar physical model is available to accurately predict future solar activity, Marshall Space Flight Center (MSFC) developed a 13-month smoothed solar flux and geomagnetic index intermediate (months) and long-range (years) statistical estimation technique¹. The reason for issuing intermediate and long-range solar activity estimates is the need for updated inputs to the upper atmosphere density models used for satellite orbital lifetime predictions and performance requirement analyses². Mission analysis and planning for future spacecraft launches and on-orbit operations require estimates of orbital lifetimes, altitudes, inclinations, and eccentricities. This report documents the MSFC 13-month smoothed solar flux and geomagnetic index intermediate and long-range statistical estimation technique, referred to as the MSFC Lagrangian Linear Regression Technique (MLLRT).

2.0 Background

Excellent surveys of various future solar activity estimation techniques were presented by Vitinskii³, Scissum⁴, Slutz, et al.⁵, and Donnelly⁶. Neural Network applications have been more recently made by Macpherson et al.⁷, and Calvo et al.⁸ and others. The linear regression method, applied by McNish and Lincoln⁹, was modified by Boykin and Richards¹⁰, and Avaritt¹¹. A later improvement applying the modified McNish and Lincoln linear regression method to develop the MSFC Lagrangian Linear Regression Technique was accomplished by Holland and Vaughan¹.

Yu. I. Vitinskii³ conducted an extensive survey and analysis of solar activity prediction methods. While recognizing the magnitude of the problem and encouraging studies of active processes taking place on the Sun to solve it, he reiterated the current status: "...we have shown that the reliability of the results obtained using these methods still leaves much to be desired." His analysis, however, showed the linear regression method usually gives accurate results to a year in advance. For several-years-in-advance, the linear regression method becomes increasingly less accurate.

McNish and Lincoln⁹ suggested that the estimation of a sunspot cycle's future behavior, based on the mean approximation of all past cycles, could be improved by adding to the mean a correction proportional to the departure of the current value of the cycle from the mean cycle. They also recommended the method not be used for making future projections longer than one year. Using a data base with two additional solar cycles, Boykin and Richards¹⁰ modified the McNish-Lincoln linear regression method so the 13-month smoothed relative sunspot number (\bar{R}) could be estimated for 10 years in advance, at quarterly rather than yearly intervals. Figure 2-1 illustrates this method that is also applicable to 13-month smoothed solar flux ($\bar{F}_{10.7}$) and geomagnetic index (\bar{A}_p).

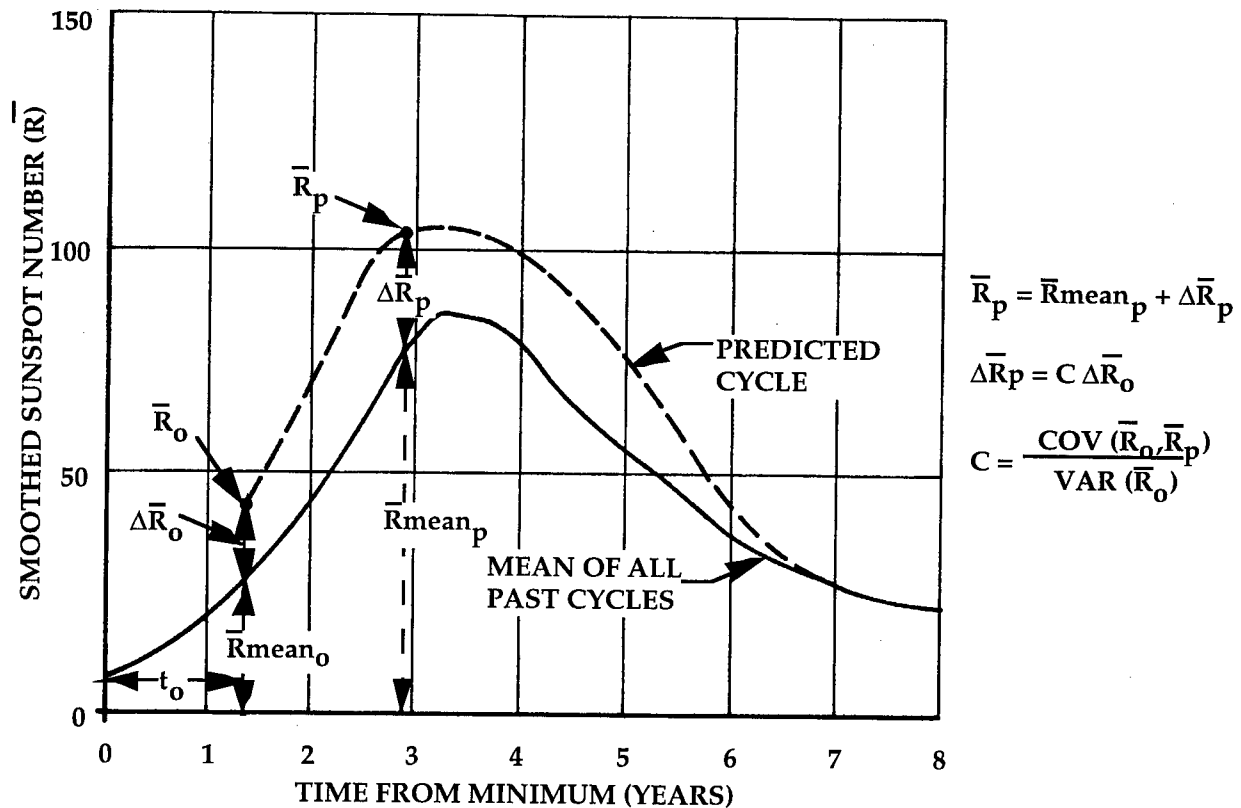


Figure 2-1. Modified McNish-Lincoln Linear Regression Method.

Holland, and Vaughan¹ determined that better statistical estimations are possible, in a chi square sense, by selecting the start and end of each set (solar cycle) at the maximum (or minimum) and normalizing the data sets using a Lagrangian linear regression statistical technique. This determination and initialization of the modified McNish-Lincoln linear regression method at the cycle's maximum or minimum establish the current MSFC statistical technique for intermediate and long-range estimation of $\bar{F}_{10.7}$ and \bar{A}_p .

3.0 MSFC Lagrangian Linear Regression Methodology

Sections 3.1 and 3.2 present the methodology used to develop the $\bar{F}_{10.7}$ and \bar{A}_p data bases and the modified McNish-Lincoln linear regression method used in the MSFC Lagrangian Linear Regression Technique (MLLRT).

3.1 Development of $\bar{F}_{10.7}$ and \bar{A}_p Data Bases

In order to use the modified McNish-Lincoln linear regression method to estimate future solar flux and geomagnetic index, $\bar{F}_{10.7}$ and \bar{A}_p , data bases were developed. Two data bases for $\bar{F}_{10.7}$ and \bar{A}_p were developed depending on what estimates are relative to time point in the present cycle. Solar cycle minimum estimates use a maximum to maximum initialized data base. Note when using the maximum to maximum data base, the cycles are identified by cycle number at maximum initialization date. For example, maximum to maximum cycle 18 starts in May 1947 with the first half of cycle 18 and the last half of cycle 19. Solar cycle maximum estimates use a minimum to minimum initialized data base. To develop a data base select the time of all maximum points of the $\bar{F}_{10.7}$ for each cycle, and use these same time

points for the \bar{A}_p , or depending on the desire estimate, select the time of all minimum points of the $\bar{F}_{10.7}$ and \bar{A}_p for each cycle. Determine all periods, P_j , related to maximum to maximum (or minimum to minimum) for the cycles. Calculate the mean period, \bar{P} , for cycles initialized at maximum (or minimum). Divide all cycle periods by \bar{P} . This gives the mean time increment, $\tau_j = P_j/\bar{P}$, to be used for the j^{th} cycle. Note that τ_j may be less than, equal to, or greater than one month. Finally, consider each cycle a group of data with m approximately equal to the number of points in \bar{P} . To obtain an $\bar{F}_{10.7}$ or \bar{A}_p value corresponding to each of the m points per cycle requires interpolation since points are not selected at precisely one month intervals for all cycles. A more detailed description of the data bases development is presented in sections 3.1.1 and 3.1.2.

3.1.1 13-Month Smoothed Solar Flux ($\bar{F}_{10.7}$) Data Base

Although some researchers believe they have sufficient reason to separate the data for sunspot cycles 1 through 8 from the total data base, the MSFC Lagrangian Linear Regression Technique for estimation of future $\bar{F}_{10.7}$ uses the observed data for all observed cycles. Including cycles 1 through 8 provides information applicable to the apparent behavior of the cycle period and to the overall magnitude during this time frame as well as a larger data base for statistical estimates. The measured $\bar{F}_{10.7}$ data base was extended back to 1749 by using Wolf's relative sunspot values, R , and a \bar{R} to $\bar{F}_{10.7}$ conversion equation. R is defined by the equation:

$$R = k(10g + f) \quad (1)$$

where R is the Wolf number, k is a correction factor to equalize counts from different observers, g is the number of groups visible on a given day, and f is the number of single spots observed on a given day¹². The R values were smoothed using the Zurich 13-month smoothing equation:

$$\bar{R}_i = \frac{1}{12} \left[\sum_{k=i-5}^{i+5} R_k + \frac{R_{i-6} + R_{i+6}}{2} \right] \quad (2)$$

where i indicates the month of interest. This smoothing technique was developed by the Swiss Federal Observatory, Zurich, Switzerland¹². Once R values are smoothed to \bar{R} values, the following equation (developed by MSFC in collaboration with Jack Slowey of the Smithsonian Astrophysical Observatory) converts recorded \bar{R} data to $\bar{F}_{10.7}$ data:

$$\bar{F}_{10.7} = 49.4 + 0.97\bar{R} + 17.6e^{(-0.035\bar{R})} \quad (3)$$

Equation (3), derived from recorded $\bar{F}_{10.7}$ and \bar{R} data from 1947 to 1978, has a linear correlation coefficient of 0.98. Figure 3-1, a plot of equation (3) with \bar{R} and $\bar{F}_{10.7}$ data to 1995, shows that equation (3) is still adequate for today's applications.

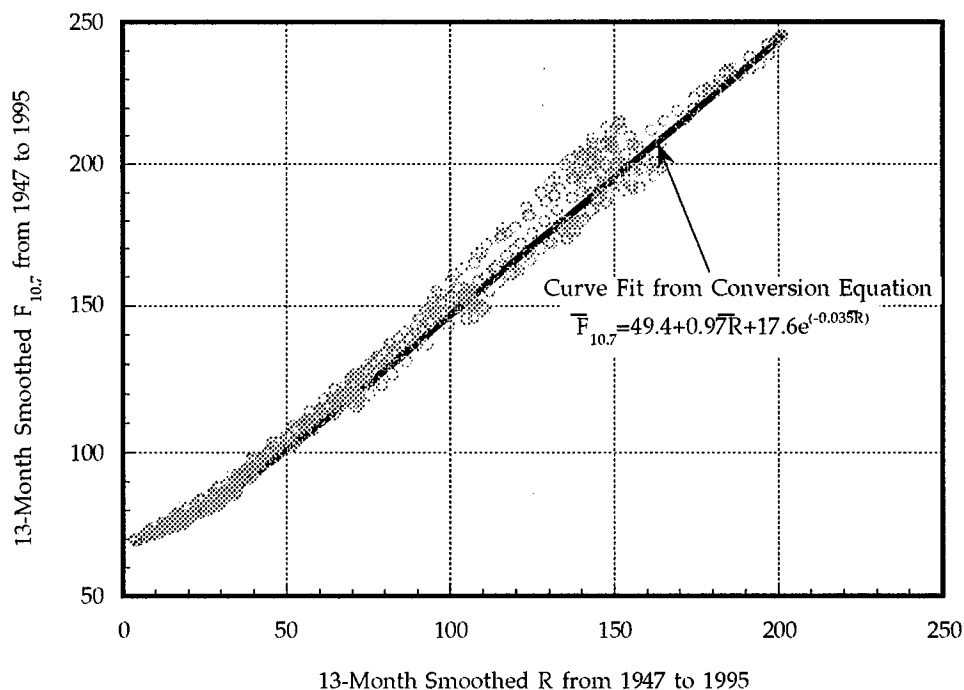


Figure 3-1. \bar{R} versus $\bar{F}_{10.7}$ with Curve Fit from Conversion Equation.

Since 1947, observed values of daily solar flux are used to compute mean monthly $F_{10.7}$ values. Use equation (2), replacing R with $F_{10.7}$, to calculate the $\bar{F}_{10.7}$. Figure 3-2 is a plot of converted and observed $\bar{F}_{10.7}$ data for all cycles since 1749.

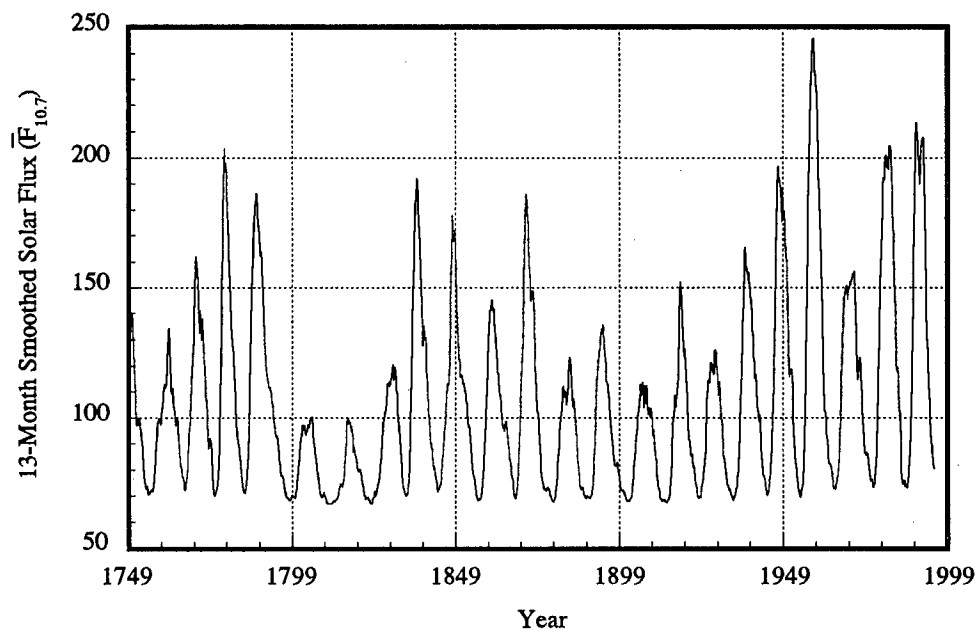


Figure 3-2. 13-Month Smoothed (Converted and Observed) Solar Flux ($\bar{F}_{10.7}$).

All solar flux cycles in the regression technique data base prior to 1947 use, as the starting point, the established \bar{R} maximum or minimum. After 1946, the maximum or minimum is based on the observed $\bar{F}_{10.7}$ maximum or minimum values computed from the

measured $F_{10.7}$ values¹. Table 3-1 has the minimum and maximum starting point years and period of each cycle with associated converted or observed $\bar{F}_{10.7}$ values. Data format is year with months in decimal form, i.e., January is 0.000 and December is 0.917. The equation to calculate the months is

$$\text{month decimal value} = (\text{month number} - 1)/12 \quad (4)$$

Table 3-1. Solar Cycle Elements Information in Data Base

Cycle	Year/Month of Minimum	Year/Month of Maximum	Minimum Solar Flux Value*	Maximum Solar Flux Value*	Rise Time in Year/Month	Fall Time in Year/Month	Min. to Min. Period (Year/Month)	Max. to Max. Period (Year/Month)
0	-	1750.250	-	139.87	-	4.833	-	11.167
1	1755.083	1761.417	70.66	134.13	6.334	5.000	11.334	8.250
2	1766.417	1769.667	72.13	162.02	3.250	5.750	9.000	8.666
3	1775.417	1778.333	70.06	203.26	2.916	6.334	9.250	9.750
4	1784.667	1788.083	71.26	186.47	3.416	10.167	13.583	17.000
5	1798.250	1805.083	68.22	100.26	6.833	5.417	12.250	11.250
6	1810.500	1816.333	67.00	99.84	5.833	7.000	12.833	13.500
7	1823.333	1829.833	67.03	120.35	6.500	4.000	10.500	7.334
8	1833.833	1837.167	70.12	191.99	3.334	6.333	9.667	10.916
9	1843.500	1848.083	71.78	177.61	4.583	7.834	12.417	12.000
10	1855.917	1860.083	68.23	144.97	4.166	7.084	11.250	10.500
11	1867.167	1870.583	69.11	185.82	3.416	8.334	11.750	13.334
12	1878.917	1883.917	67.85	123.04	5.000	6.250	11.250	10.083
13	1890.167	1894.000	69.01	135.50	3.833	8.000	11.833	12.083
14	1902.000	1906.083	68.02	113.54	4.083	7.417	11.500	11.500
15	1913.500	1917.583	67.54	152.07	4.083	6.000	10.083	10.667
16	1923.583	1928.250	69.30	126.32	4.667	5.417	10.084	9.000
17	1933.667	1937.250	68.36	165.30	3.583	6.833	10.416	10.083
18	1944.083	1947.333	70.33	214.39	3.250	6.917	10.167	10.834
19	1954.250	1958.167	69.69	245.45	3.917	6.583	10.500	12.333
20	1964.750	1970.500	72.59	156.34	5.750	5.917	11.667	10.833
21	1976.417	1981.333	73.27	204.55	4.916	5.334	10.250	8.084
22	1986.667	1989.417	72.90	213.11	2.750			
Summary:								
Median (50%)			69.50	156.34	4.08	6.33	11.25	10.83
Mean			69.75	160.70	4.38	6.49	11.03	10.87
Standard Deviation (σ)			1.93	39.92	1.21	1.36	1.19	2.12

* Converted prior to 1947 and observed after 1946

The converted and observed $\bar{F}_{10.7}$ data are Lagrangian interpolated to normalize the data for the 132 months from the maximum or minimum cycle starting dates. The data is stored by month and cycle number to construct a data base for use in the modified McNish-Lincoln linear regression method.

3.1.2 13-Month Smoothed Geomagnetic Index (\bar{A}_p) Data Base

Because the measured geomagnetic index (A_p) data base is relatively short (1932 to 1996), it was extended back to 1884 using mean monthly magnetic character figure (C_i) data. The mean monthly C_i data is converted to 13-month smoothed data using equation (2) and replacing R with C_i . Once the mean monthly C_i is converted to 13-month smoothed magnetic character figure (\bar{C}_i), use equation (5) to convert the extended record of \bar{C}_i data to \bar{A}_p values.

$$\bar{A}_p = 2.8068 e^{2.393 \bar{C}_i} \quad (5)$$

This equation, derived from recorded¹³ C_i and A_p data from 1932 to 1963, has a linear correlation coefficient of 0.80. Figure 3-3 is a plot of equation (5) with \bar{C}_i and \bar{A}_p from 1932 to 1963.

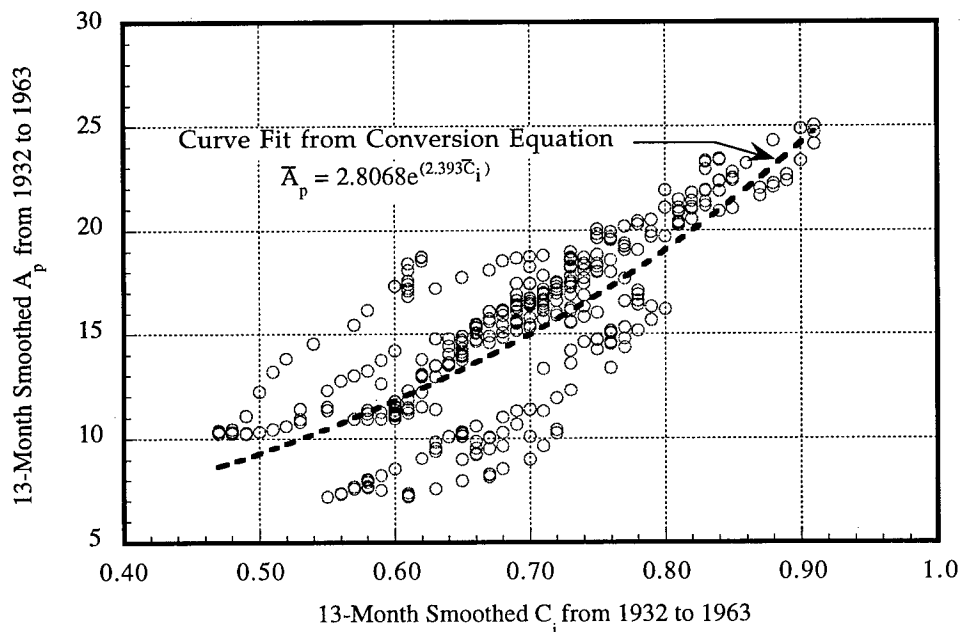


Figure 3-3. \bar{C}_i versus \bar{A}_p with Curve Fit from Conversion Equation.

After 1931, the measured values of daily A_p are used to compute the mean monthly value. Use equation (2), replacing R with A_p , to calculate \bar{A}_p . Figure 3-4 is a plot of converted and recorded \bar{A}_p data for all cycles since 1884.

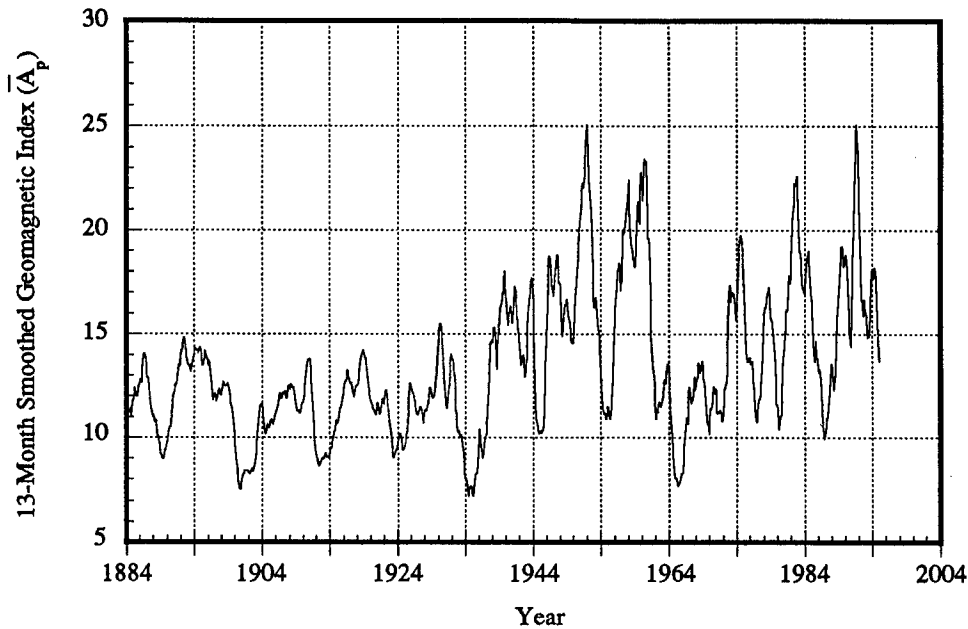


Figure 3-4. 13-Month Smoothed (Converted and Recorded) Geomagnetic Index (\bar{A}_p).

The converted and recorded \bar{A}_p data are Lagrangian interpolated to normalize the data for the 132 months from the maximum or minimum cycle starting dates. The data is stored by month and cycle number to construct a data base for use in the modified McNish-Lincoln linear regression method.

3.2 Modified McNish-Lincoln Linear Regression Method

The MLLRT uses the Boykin and Richards¹⁰ modified McNish-Lincoln linear regression method (for the linear least squares method development see appendix A) and an appropriately constructed data base that starts at the maximum or minimum to estimate the balance of the present cycle where the cycle is defined from minimum to minimum or maximum to maximum. This method is summarized in the following steps:

1. Mean $\bar{F}_{10.7}$ or \bar{A}_p is calculated from the completed cycles in the $\bar{F}_{10.7}$ or \bar{A}_p data base constructed using the Lagrangian interpolated data points for use in the McNish-Lincoln linear regression method. This mean also estimates $\bar{F}_{10.7}$ or \bar{A}_p for the next cycle with \bar{P} .
2. McNish-Lincoln linear regression method produces a statistical estimate for the rest of the present cycle using one linear coefficient. The period for the present cycle, for which estimates of solar activity are being calculated, is the \bar{P} .
3. Since, for the present cycle, only 21 or 22 corresponding points are available for a linear regression fit of the estimated point to the last observed point, to justify calculating a standard deviation based on a normal distribution function is difficult. This non-normal distribution function produces upper and lower bounds that can and do go below the parameter physical limits. Despite being a non-normal distribution, the data are standardized to make calculations easier. The actual distribution of deviations from the smoothed linear regression line and mean line is divided by the standard deviation and used to determine the upper and lower bounds at predetermined percentile levels. Upper and lower bounds are

calculated by the Quantile method. The equation used is:

$$Q(x_i) = i/(n+1) \quad (6)$$

where Q is quantile, i equals 1 through the total number of completed cycles, and n is equal to the total number of completed cycles. Once the quantile is calculated the percentile is:

$$\text{Percentile}(y) = 100.0 Q(x) \quad (7)$$

Rationale for using the Quantile method is discussed in appendix C.

4. Between the upper and lower bounds discussed in step 3 is the "error space" on a two-dimensional plot of $\bar{F}_{10.7}$ or \bar{A}_p versus time, t.

4.0 Application of MSFC Lagrangian Linear Regression Technique

The computer programs to calculate statistical estimates for future $\bar{F}_{10.7}$ and \bar{A}_p cycles and to perform subsequent statistical analyses were developed in FORTRAN. The following sections describe how the programs implement the MLLRT. MLLRT consists of three main programs: (1) the $\bar{F}_{10.7}$ and \bar{A}_p preprocessor computer program (PREPROC), (2) modified McNish-Lincoln linear regression computer program (FORECAST), and (3) statistical estimate output report computer program (REPORT). These programs are used to produce the MSFC monthly solar activity memorandum data (see appendix E). Figure 4-1 is a summary block diagram showing the flow of files through these programs.

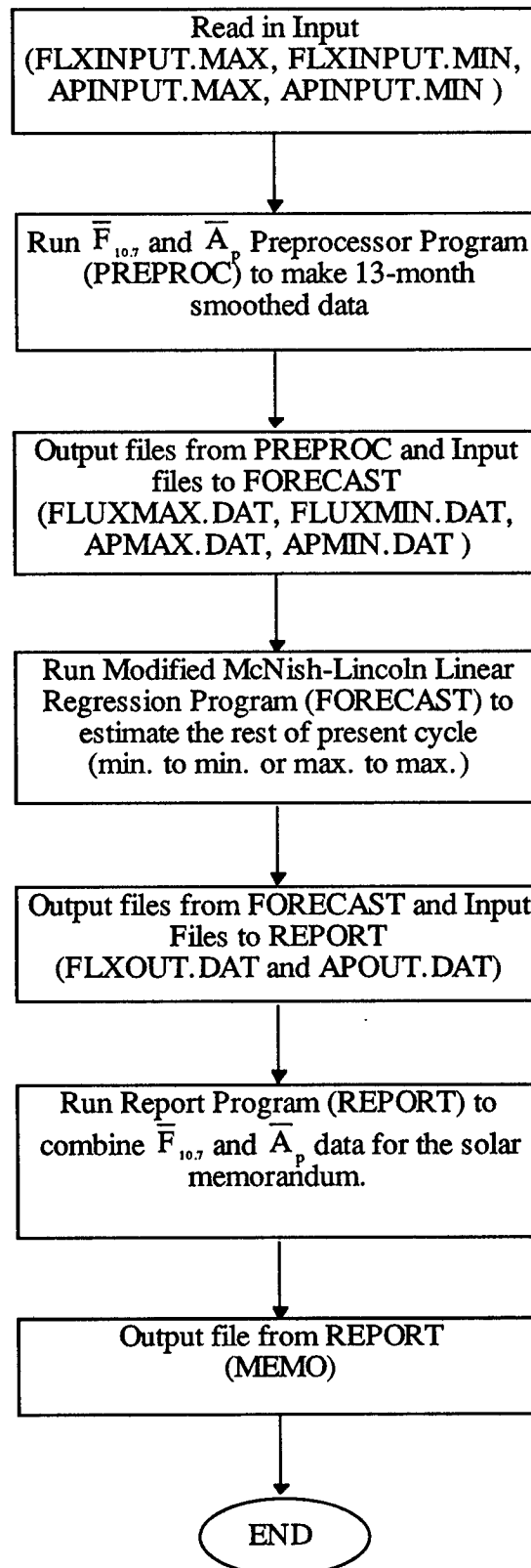


Figure 4-1. MSFC Lagrangian Linear Regression Technique Block Diagram.

4.1 $\bar{F}_{10.7}$ and \bar{A}_p Preprocessor Computer Program

PREPROC uses mean monthly R , $F_{10.7}$, C_i , and A_p data to supply the linear regression program with 13-month smoothed data initialized at the cycle's maximum or minimum solar activity dates. Table 3-1 summarizes the initialization $\bar{F}_{10.7}$ dates for present and past solar cycles. Data are initialized at the $\bar{F}_{10.7}$ maximum (minimum) dates to enable the program to better estimate duration of the present cycle and when its minimum (maximum) will occur. Once the minimum (maximum) is determined, the preprocessor supplies the linear regression computer program with 13-month smoothed data initialized at the minimum (maximum) solar activity date for the computer program to provide a better estimate for the "new" present cycle and occurrence of its maximum (minimum). Steps in section 3.1 are the methodology for the preprocessor computer program. How individual files are generated for FORECAST is given in the next paragraph.

PREPROC starts with the monthly mean values of R and $F_{10.7}$ input data (flxinput.max or flxinput.min) or the C_i and A_p input data (apinput.max or apinput.min). The flxinput.max or flxinput.min data have two sources: the monthly mean R data from M. Waldmeier's book¹², and the National Research Council of Canada monthly mean $F_{10.7}$ data. Both sets of data are smoothed by equation (2) in section 3.1.1 and \bar{R} data are converted to $\bar{F}_{10.7}$ by equation (3). The apinput.max or apinput.min data have two sources: the monthly mean geomagnetic C_i data from the Handbook of Geophysics and Space Environments¹³ and the Institute for Geophysics in Gottingen, Germany monthly mean A_p data. Both sets of data are smoothed by equation (2) in section 3.1.1 and \bar{C}_i data are converted to \bar{A}_p data by equation (5). Once the $\bar{F}_{10.7}$ or \bar{A}_p sets are complete, the data are grouped into cycles then normalized to 132 points by a 3-point Lagrange interpolation subroutine (COMB_YLGINT). $\bar{F}_{10.7}$ or \bar{A}_p output from the preprocessor program is used in a create file subroutine (CREAT_BLOCK) which formats the normalized data into a new output file (fluxmax.dat or fluxmin.dat) or (apmax.dat or apmin.dat) for FORECAST. Figure 4-2 is a summary block diagram showing the flow of files through PREPROC and its subroutines to create the input file for FORECAST.

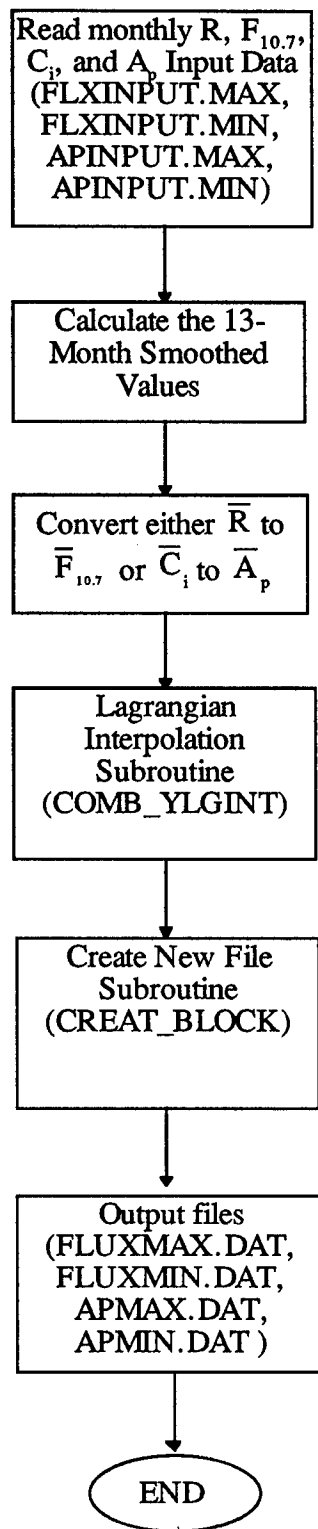


Figure 4-2. $\bar{F}_{10.7}$ and \bar{A}_p Preprocessor Computer Program Block Diagram.

4.2 Modified McNish-Lincoln Linear Regression Computer Program

FORECAST estimates $\bar{F}_{10.7}$ and \bar{A}_p for the rest of the minimum to minimum or maximum to maximum present cycle using the methodology discussed in section 3.2. A description of generating individual files via FORECAST follows.

FORECAST reads the $\bar{F}_{10.7}$ or \bar{A}_p data (fluxmax.dat, fluxmin.dat or apmax.dat, apmin.dat) and standardizes completed cycles data. $\bar{F}_{10.7}$ or \bar{A}_p data enter the modified McNish-Lincoln subroutine (LINCMC) which first calculates the mean of the completed cycles. LINCMC calculates next the difference between the actual and mean used in calculating the variance matrix (A) and covariance matrix (B). LINCMC uses an inverse matrix subroutine (GJR) to calculate the A^{-1} matrix and subsequently the regression coefficient matrix (C). Using the C matrix LINCMC estimates the rest of the present cycle (step 2 section 3.2). Using the historical data base, the next cycle is estimated (step 1 section 3.2). Percentiles subroutine (PRCNTILE) calculates the 95.0 and 5.0 percentile values for the present cycle. PRCNTILE divides the difference data by the standard deviation then calculates the desired quantile value by the quantile subroutine (QUANTILE). QUANTILE arranges the data in ascending order in subroutine (RANKIT) then calculates the quantile values (equation 6 in section 3.2 step 3). PRCNTILE enters the find-the-percentile-subroutine (FNDPRCNT) and calculates the needed percentile value. FNDPRCNT determines the location of the desired quantile value then uses a linear interpolation subroutine (INTERP) if the value is not the exact quantile calculated by QUANTILE. Once the quantile values are determined, each is multiplied by 100 for the percentile value. FORECAST next enters a smoothing subroutine (PCHMAX or PCHMIN). The find-cubic-inflection-point subroutine (INFLBS) locates the cubic inflection point between the McNish-Lincoln estimate and the best statistical estimate. The percentiles have a cubic inflection point also. The cubic curve fit subroutine (CUBFIT) smoothes the curve between the two estimates (see appendix B for cubic connection theory). Once smoothed, the final combined data are formatted for plots and tables. Figure 4-3 is the modified McNish-Lincoln linear regression computer program block diagram. Appendix D is a listing of LINCMC and QUANTILE subroutines.

ERRATA

NASA TECHNICAL MEMORANDUM 4759

STATISTICAL TECHNIQUE FOR INTERMEDIATE AND LONG-RANGE ESTIMATION OF 13-MONTH SMOOTHED SOLAR FLUX AND GEOMAGNETIC INDEX

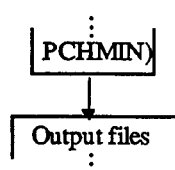
by K.O. Niehuss, H.C. Euler, Jr., W.W. Vaughan

September 1996

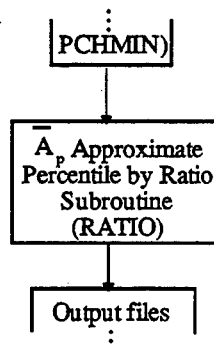
1. Page 12, second paragraph, ninth line:
Change ...calculates the 95.0 and 5.0 percentile...
To read ...calculates the desired percentile...
2. Page 12, second paragraph, twenty second line:
Change...plots and tables. Figure 4-3 is...
To read ...plots and tables. However, before the \bar{A}_p data are formatted for plots and tables, the \bar{A}_p approximate percentile by ratio subroutine (RATIO) approximates the desired percentile values if \bar{A}_p is above or below the possible calculated percentile value by the quantile method. Figure 4-3 is...

3. Page 13, Figure 4-3, first column:

Change



To read



4. Page 14, last paragraph:
Change Appendix E ... \bar{A}_p estimates.
To read Appendix E discussion above applies to the \bar{A}_p estimates except for the percentile values. Since the \bar{A}_p number of cycles is smaller than the $\bar{F}_{10.7}$ number of cycles (10 cycles instead of 22), the upper and lower bounds will be calculated at 90 and 10 percentile instead of 95 and 5 percentile. To give the lifetime prediction models consistent input data, and to compensate for the small \bar{A}_p data base, an approximated 95 and 5 percentile was calculated using a ratio technique. The computer program calculates the 5 percent value from the 80 percent value range using a ratio of the data. Then adding and subtracting the 5 percent value to and from the 90 and 10 percentiles gives the approximated 95 and 5 percentiles. The approximated percentiles for the \bar{A}_p make a very small difference in the lifetime calculations from upper atmosphere model orbital density values and provide \bar{A}_p percentiles inputs consistent with the $\bar{F}_{10.7}$ percentiles.
5. Page D-6, DO Loop of Quantile Subroutine, twenty-second through twenty-fifth lines:
Change DO I=1,NU2
:
END DO

To read QUANTNUM = NU2-NU1+1
DO I=NU1,NU2
QUANT(I)=FLOAT(I+1-NU1)/(FLOAT(QUANTNUM)+1.0)
END DO
6. Page E-2, twenty-third line:
After ...values.
Add the sentence... Due to the small sample size, the 95 percentile and 5 percentile values for the geomagnetic index are approximated.

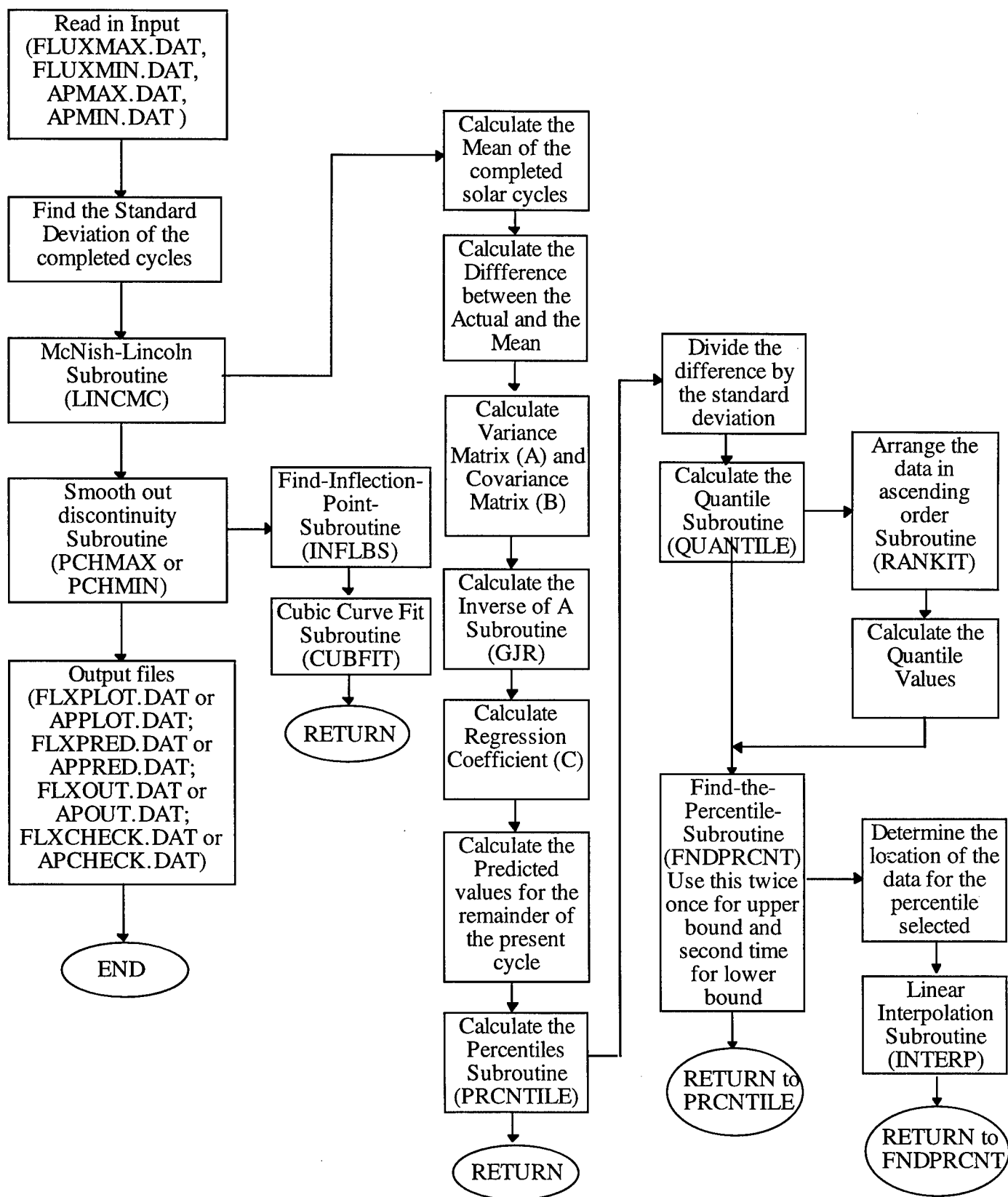


Figure 4-3. Modified McNish-Lincoln Linear Regression Computer Program Block Diagram.

4.3 Statistical Estimation Output Report Computer Program

REPORT combines the $\bar{F}_{10.7}$ and \bar{A}_p data, reads the FORECAST output files (fluxout.dat and apoutput.dat), and merges the data via the merge subroutine (MERGE). The memorandum printout subroutine (FRMEMO) takes the data from MERGE and prints the table for the MSFC monthly solar activity memorandum. Figure 4-4 is a summary block diagram of the report program; appendix E is an example of the memorandum.

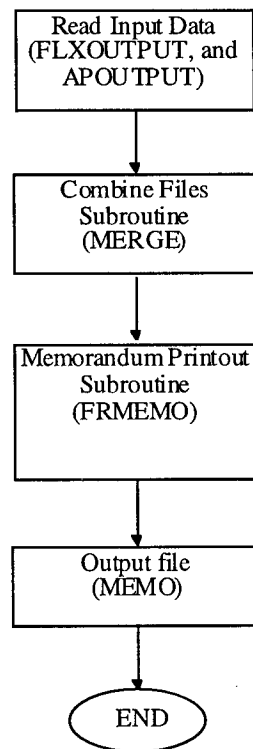


Figure 4-4. Report Program Block Diagram.

Appendix E provides a sample of the MLLRT computer program output during a maximum to maximum cycle. The present 132-month period cycle initialization month is June 1989 (cycle 22 maximum) and the $\bar{F}_{10.7}$ value is 213.1. Using the statistical Lagrangian data base for the converted and observed smoothed solar flux, the estimate of the 95, ~50, and 5 percentile values for $\bar{F}_{10.7}$ is arranged in a monthly sequence through the balance of cycle 22, through cycle 23, and into the first half of cycle 24. Obtainable from the 95, ~50, and 5 percentile $\bar{F}_{10.7}$ values for the present cycle is an estimate of the range for the upcoming solar flux cycle maximum values. The range estimate given for the 132 months from the maximum of cycle 23 to the maximum of cycle 24 is determined from the mean (~50%) cycle and the upper (95%) and lower (5%) bounds based on the converted and observed $\bar{F}_{10.7}$ data base (currently 21 cycles). The tie-in for the present to next cycle uses a cubic connection from the nearest inflection points on the rising leg of cycle 23 to the maximum of the 95, ~50, 5 percentile values of the flux for all cycles. This defines the maximum of cycle 23 and the extension of the estimate into the next cycle. Tie-ins accomplished with a cubic curve fit explain the relatively smoothed appearance of the curve in this tie-in area. The memorandum in appendix E, presents the 95 and 5 percentile values for the past 21 cycles as a matter of reference relative to the mean (~50%) period used in the statistical estimation computer program.

Appendix E discussion above also applies to the \bar{A}_p estimates.

5.0 Engineering Use of MSFC Solar Activity Statistical Estimation Technique Products

MLLRT's future $\bar{F}_{10.7}$ and \bar{A}_p estimates are intended mainly for use as inputs to upper atmosphere models, i.e., Marshall Engineering Thermosphere Model, (NASA CR-179359 and CR-179389), NASA/MSFC Global Reference Atmosphere Model, (NASA TM-4715), and others. Figure 5-1, for example, depicts inputs related to the MSFC spacecraft orbital lifetime model.

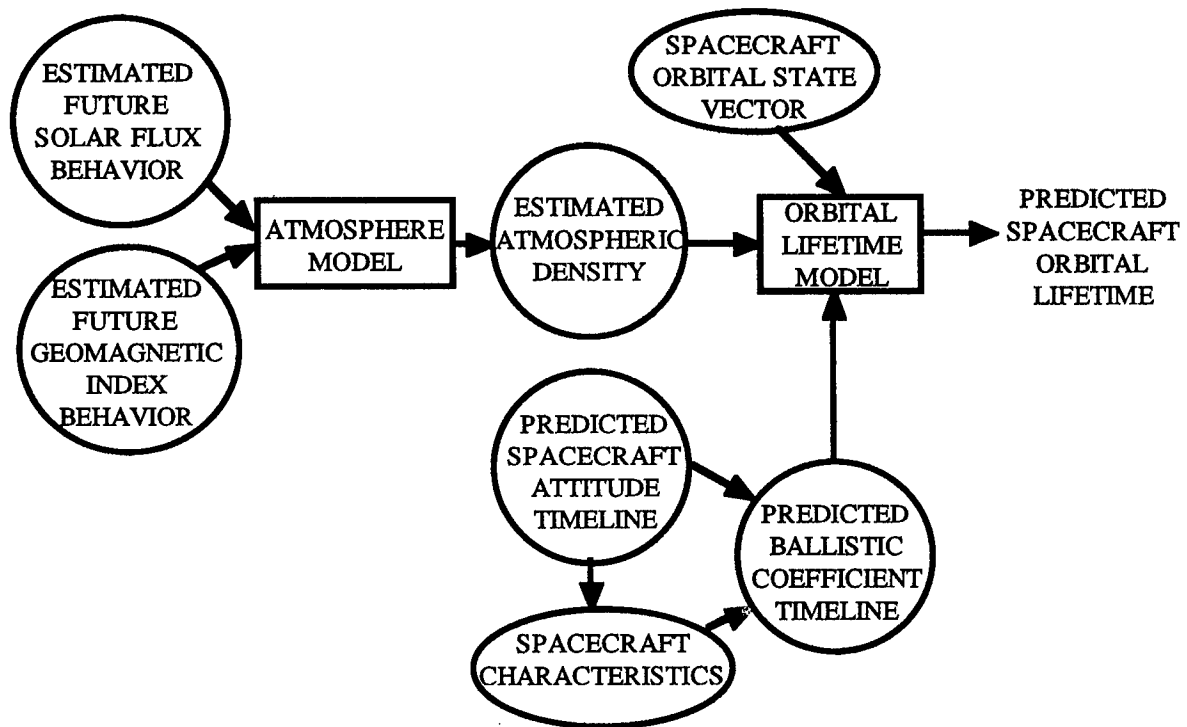


Figure 5-1. MSFC Spacecraft Orbital Lifetime Prediction Model Block Diagram^{1,14}.

Since there is no method for intermediate and long-term predictions of daily $F_{10.7}$ and a_p , orbital lifetime models use the $\bar{F}_{10.7}$ and \bar{A}_p estimates². Orbital lifetime estimates, control analysis programs, et al. , require a specific date to associate with the future estimate of $\bar{F}_{10.7}$ and \bar{A}_p to compute corresponding atmospheric density. Future estimated $\bar{F}_{10.7}$ and \bar{A}_p data points should be identified with the first day of the given month.

For spacecraft projects requiring a minimum risk design lifetime orbital altitude(s) and a specified control capability, the 95 percentile estimates of $\bar{F}_{10.7}$ and \bar{A}_p are recommended. Taking into account the short-term (days) dynamics, these estimates permit the design of a statistically conservative spacecraft lifetime and control capability at a given orbital altitude(s). The lifetime determination should be based on the most current intermediate and long-range statistical estimate of the future solar flux and geomagnetic index consistent with the critical project development decision time points prior to planned launch of the spacecraft.

Changes in orbital density associated with short-term variations in the daily $F_{10.7}$ and a_p required as inputs to the upper atmospheric models, are not represented by the $\bar{F}_{10.7}$ and \bar{A}_p statistical estimates given in the MSFC monthly solar activity memorandum. Future changes in total atmospheric density cannot be estimated with any acceptable degree of statistical confidence

using existing techniques. Representative data sets, based on past daily $F_{10.7}$ and a_p values, may be utilized to compute this dynamic component of the orbital altitude density.

The design requirements for solar activity and associated on-orbit density values are specified in the appropriate spacecraft and space vehicle project design requirements documentation. Consult these documents for this information.

6.0 Concluding Remarks

Because no physical solar activity prediction models are accepted generally by the scientific community, MSFC developed the MLLRT. MLLRT provides $\bar{F}_{10.7}$ and \bar{A}_p values for input parameters to the MSFC orbital altitude neutral atmosphere models noted in section 5.0. Although the atmosphere models are built on inputs of $F_{10.7}$ and a_p , the $\bar{F}_{10.7}$ and \bar{A}_p values are the lowest level of smoothing with any acceptable level of confidence that lends itself to reasonably accurate statistical prediction.

The technique utilized by MSFC is based on a small sample size and a nongaussian distribution of $\bar{F}_{10.7}$ and \bar{A}_p . To estimate future activity MSFC uses a mean cycle and deviations derived from previous cycles, combined with a one-term linear regression, to estimate future activity of the present cycle. Since intermediate term (months) estimates are more accurate than long range (years) estimates, the data base is initialized at both established maximum and minimum values of solar cycle activity. Based on data from previous cycles, the probability is 90 percent that the actual future $\bar{F}_{10.7}$ or \bar{A}_p values will be within the current MLLRT computer program output for estimated 95 and 5 percentile values. The computer program may be adjusted to accommodate any desired percentiles between 95 and 5.

To provide an assessment of the MLLRT computer program's products, appendix F presents plots (Figure F-1 through F-6) of the $\bar{F}_{10.7}$ future estimates from the MLLRT for a 5-year period from date of estimation for several different solar cycles. Three of the solar flux estimation periods are for 11-year periods and began at the minimum between cycles 19-20, 20-21, and 21-22 and three at the maximum of cycles 20, 21 and 22.

7.0 References

1. Holland, R. L., and William W. Vaughan: Lagrangian Least-Squares Prediction of Solar Flux ($\bar{F}_{10.7}$). *Journal of Geophysical Research*, Vol. 89, No. A1, pp. 11-16, January 1, 1984.
2. Dreher, P. E., and A. T. Lyons: Long-Term Orbital Lifetime Predictions. NASA Technical Paper 3058, George C. Marshall Space Flight Center, Huntsville, Alabama, October 1990.
3. Vitinskii, Yu, I.: Solar-Activity Forecasting. Translated from Russian, Israel Program for Scientific Translations, Jerusalem. NASA TT F-289, Clearing House for Federal Scientific and Technical Information, Springfield, Virginia, 1965.
4. Scissum, J.: Survey of Solar Cycle Prediction Models. NASA TM-53593, 1967.
5. Slutz, R. J., Gray, T. B., West, M. L., Stewart, F. G., and Leftin, Margo: Solar Activity Prediction. NASA CR-1939, November 1971.
6. Donnelly, R. F. (Ed.): Solar-Terrestrial Predictions Proceedings, vols. 1, 2, 3, and 4, U. S. Government Printing Office, Washington D. C., 1979.
7. Macpherson, K.P., A. J. Corwayan J.C. Brown: Prediction of Solar and Geomagnetic Activity Data Using Neural Networks. *Journal of Geophysical Research*, Vol. 100, No. A11, pp. 21,735-21,744, November 1, 1995.
8. Calvo, R. A. , H. A. Ceecalto, and R. D. Piacentini: Neural Network Prediction of Solar Activity. *AstrophysicJournal*, Vol. 444, pp. 916-920, May 10, 1995.
9. McNish, A. G., and Lincoln, J. V.: Prediction of Sunspot Numbers. *Transactions, American Geophysical Union*, 30, 1949, p. 673.
10. Boykin, E. P., and Richards, T. J.: Application of the Lincoln McNish Technique to the Prediction of the Remainder of the Twentieth Sunspot Cycle. Lockheed Missiles and Space Company, Technical Memorandum 54/30-89, Huntsville, Alabama, March 1966.
11. Avaritt, D.: Sunspot Prediction Program (Sunspot Deck). Program Document 3F0360, Computer Sciences Corporation, Huntsville, Alabama, 1972.
12. Waldmeier, M.: *The Sunspot Activity in the Years 1610-1960*. Zurich Schulthess and Company, Switzerland, 1961.
13. Valley, Shea L. (Ed.): *Handbook of Geophysics and Space Environments*, McGraw-Hill Book Company, New York, April 1965.
14. Dreher, P. E., Little, R. P., and Wittenstein, G.: Skylab Orbital Lifetime Prediction and Decay Analysis. NASA, Technical Memorandum 78308, Huntsville, Alabama, November 1980.

8.0 Bibliography

Bumba, V. and J. Uleczeck (Ed.): *Basic Mechanism of Solar Activity*. D. Reidel Publishing Company, Boston, MA, ISBN#90-277-0680-8, 1976.

Eckstein, Herbert P.: *Solar Activity: A Bibliography*. RSIC-706, Redstone Scientific Information Center, U.S. Army Missile Command, September 1967.

Jordan, Stuart (Ed.): *The Sun as a Star*. NASA SP-450, 1981.

Mayland, P.N.: *Derivation, Meaning, and Use of Geomagnetic Indices*. American Geophysical Union, Washington, D.C., ISBN#0.87590-022-4, 1980.

Rubashev, B.M. : *Problems of Solar Activity*. NASA TT F-244, December 1964.

Schove, D. Justin (Ed.): *Sunspot Cycles*. Hutchinson Ross Publishing Company, Stroudsburg, Penn, ISBN# 0-87933-424-X, 1983.

Vorchishin, K. S. and M. I. Stodilko: *The Statistical Stability of Solar Cycling*. *Astron. J.*, Vol. 59, No. 6, pp. 1171-1183, 1983.

Wilson, Robert M.: *Sunspot Variation and Selected Associated Phenomena: A Look at Cycle 21 and Beyond*. NASA TM-82474, February 1982.

Wilson, Robert M.: *A Comparative Look at Sunspot Cycles*. NASA TP-2325, May 1984.

Xanthakis, J. (Ed): *Solar Activity and Related Interplanetary and Terrestrial Phenomena*. Springer-Verlag Publishers, New York, New York, ISBN#3-540-06314-5, 1973.

APPENDIX A Linear Least Squares Model Development

Robert L. Holland
Computer Sciences Corporation
Huntsville, Alabama

From the Lagrangian interpolated cycles, there are currently 21 complete cycles from maximum to maximum peaks. There are 132 interpolated values in each of these cycles. In the following development of the modified McNish-Lincoln linear regression method each point in the data will be referred to as R_{ij} where i refers to the i^{th} point within a cycle (1 to 132) and j refers to the cycle number (1 to 22). Other variables will be defined as they appear.

The expected value of R' at p -points beyond m , where m is the number of observed values in the current cycle can be written as a linear combination of any number $k \leq m$ coefficients times the deviations from the means of the k -known 13-month smoothed monthly values within the current cycle.

Holland, Rhodes, and Euler ^{A-1} included the time derivatives in the linear combination model. In the following development, the derivatives will be omitted since their results showed that including more than one coefficient and the derivative did not improve the confidence level nor the goodness of fit. For $k=2$, the model requires 2 coefficients and 2 independent variables and likewise for $k \leq m$. The $k = 2$ model will be developed for simplicity then specialized for $k = 1$, which is the MLLRT model in use at MSFC. The model may be presented classically as

$$z' = ax + by \quad (\text{A-1})$$

where x and y are the two independent variables, z' is the dependent variable, and a and b are the linearly-related (regression) coefficients.

z' is the model predicted value of the actual z data with N known (x_j, y_j, z_j) corresponding points. It is required to find the coefficients a and b which best fit the data in a least squares to minimize the sum of the squares of the deviations of the chosen model (A-1) predictions from the data. Substituting the data points x_j, y_j in the right side of equation (A-1) gives a model predicted z'_j , i.e.,

$$z'_j = ax_j + by_j, \quad (\text{A-2})$$

$1 \leq j \leq N$ (the number of known data points).

Now it is required to minimize the sum of the squared differences (deviations) of these predicted z_j' from the data z_j . These deviations d_j are

$$d_j = z_j - z_j' = z_j - (ax_j + by_j) \quad (\text{A-3})$$

There are N of these deviations for each corresponding point within the 132-point Lagrangian interpolated cycle. Denoting d_j, x_j, y_j, z_j and z_j' as n -dimensional vectors defined as

$$\mathbf{d} \equiv \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{pmatrix} \quad \mathbf{z} \equiv \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_N \end{pmatrix} \quad \mathbf{z}' \equiv \begin{pmatrix} z_1' \\ z_2' \\ \vdots \\ z_N' \end{pmatrix} \quad \mathbf{x} \equiv \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} \quad \mathbf{y} \equiv \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}$$

then equation (A-3) can be written in vector-matrix notation as

$$\mathbf{d} = \mathbf{z} - (\mathbf{ax} + \mathbf{by}) \quad (\text{A-4})$$

The sum of the squares of the deviations of the data (\mathbf{z}) from the model prediction ($\mathbf{ax} + \mathbf{by}$) are obtained by taking the dot product of \mathbf{d} with itself. This gives

$$d^2 \equiv \mathbf{d} \cdot \mathbf{d} = [\mathbf{z} - (\mathbf{ax} + \mathbf{by})] \cdot [\mathbf{z} - (\mathbf{ax} + \mathbf{by})] \quad (\text{A-5})$$

carrying our this dot product gives

$$\begin{aligned} d^2 &\equiv \mathbf{z} \cdot \mathbf{z} - 2\mathbf{z} \cdot (\mathbf{ax} + \mathbf{by}) + (\mathbf{ax} + \mathbf{by}) \cdot (\mathbf{ax} + \mathbf{by}) \\ &= \mathbf{z}^2 - 2a\mathbf{z} \cdot \mathbf{x} - 2b\mathbf{z} \cdot \mathbf{y} + (\mathbf{ax} + \mathbf{by})^2 \end{aligned}$$

or

$$d^2 = z^2 - 2az \cdot x - 2bz \cdot y + a^2x^2 + 2abx \cdot y + b^2y^2 \quad (A-6)$$

To minimize d^2 , the partial derivatives of d^2 with respect to a and b must be zero. Performing these derivatives gives

$$\frac{\partial d^2}{\partial a} = -2z \cdot x + 2ax^2 + 2bx \cdot y = 0 \quad (A-7)$$

$$\frac{\partial d^2}{\partial b} = -2z \cdot y + 2by^2 + 2ax \cdot y = 0 \quad (A-8)$$

Rearranging terms and dropping the 2's gives

$$ax^2 + bx \cdot y = z \cdot x \quad (A-9)$$

$$ax \cdot y + by^2 = z \cdot y \quad (A-10)$$

or in the adopted vector-matrix notation this becomes

$$\begin{array}{ccc} \begin{pmatrix} z \cdot x \\ z \cdot y \end{pmatrix} & = & \begin{pmatrix} x^2 & x \cdot y \\ x \cdot y & y^2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} \\ \downarrow & & \downarrow \quad \downarrow \\ \mathbf{v} & & (\mathbf{M}) \quad \mathbf{c} \end{array} \quad (A-11)$$

Thus from the vector-matrix equation

$$\mathbf{v} = (\mathbf{M}) \mathbf{c} \quad (A-12)$$

the \mathbf{c} coefficient vector is determined by multiplying equation (A-12) through by the inverse of (\mathbf{M}) , i.e.,

$$(\mathbf{M})^{-1} \mathbf{v} = (\mathbf{M})^{-1}(\mathbf{M})\mathbf{c} \Rightarrow \mathbf{c} = (\mathbf{M})^{-1} \mathbf{v} \quad (A-13)$$

notice that $(\mathbf{M})^{-1}$ and \mathbf{v} on the right are all known data.

For a single independent variable x , then $b=0$ and there is no y variable. In this special one coefficient case, equation (A-11) becomes

$$\mathbf{z} \bullet \mathbf{x} = a\mathbf{x}^2 \quad \text{or} \quad a = \frac{\mathbf{z} \bullet \mathbf{x}}{\mathbf{x}^2} \quad (\text{A-14})$$

It now remains to associate the foregoing with the actual solar flux $F_{10.7}$ and geomagnetic index A_p data and the statistical method (MLLRT) currently in use at MSFC.

$$\mathbf{x} \rightarrow \Delta \mathbf{R}_m, \Delta \mathbf{R}_m = \mathbf{R}_m - \bar{\mathbf{R}}_m,$$

$$\mathbf{y} \rightarrow \Delta \mathbf{R}_{m-1}, \Delta \mathbf{R}_{m-1} = \mathbf{R}_{m-1} - \bar{\mathbf{R}}_{m-1},$$

$$\text{and} \quad \bar{\mathbf{R}}_i = \frac{1}{N} \sum_{j=1}^N \mathbf{R}_{ij} \quad 1 \leq i \leq 132$$

$$\text{also} \quad \mathbf{R}_m \equiv \begin{pmatrix} \mathbf{R}_{m1} \\ \mathbf{R}_{m2} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{R}_{mN} \end{pmatrix} \quad \Delta \mathbf{R}_m \equiv \begin{pmatrix} \Delta \mathbf{R}_{m1} \\ \Delta \mathbf{R}_{m2} \\ \cdot \\ \cdot \\ \cdot \\ \Delta \mathbf{R}_{mN} \end{pmatrix}$$

m is the most current 13-month smoothed month number from the beginning of the current cycle. $a \rightarrow c$ and $b \rightarrow o$ are the coefficients corresponding with c in the MLLRT model which is a one coefficient model. \mathbf{z} is any 21-point data vector from the 132 in the cycle, excluding $\Delta \mathbf{R}_m$ and $\Delta \mathbf{R}_{m-1}$. \mathbf{z} is the model prediction for any of the known points within any of the known cycles or for the points in the current cycle.

With the above association, the MLLRT model can now be presented as

$$\Delta \mathbf{R}'_{m+p} = c_{m+p} \Delta \mathbf{R}_m \quad (\text{A-15})$$

where

$$c = \frac{\Delta \mathbf{R}'_m \bullet \Delta \mathbf{R}_m}{\Delta \mathbf{R}_m \bullet \Delta \mathbf{R}_m}$$

m is the number of known points within the current cycle. p is any point beyond m within the 132-point cycle. Since the motivation in the MLLRT model is for projecting beyond the N known data points and beyond m in the current cycle, then $p > m$ for the $N+1$ current cycle.

From equation (A-15) the MLLRT one coefficient model is

$$R_{m+p, N+1} = \bar{R}_{m+p} + c_{m+p} \Delta R_{m, N+1} \quad (A-16)$$

where

$$\bar{R}_{m+p} = \frac{1}{N} \sum_{j=1}^N R_{m+p, j}$$

and

$$\Delta R_m = R_{m, N+1} - \bar{R}_m$$

Reference

- A-1. Holland, R. L., C. A. Rhodes, and H. C. Euler: Lagrangian Least Squares Prediction of Solar Activity. NASA TM-82462, George C. Marshall Space Flight Center, Huntsville, Alabama, April 1982.

APPENDIX B

Cubic Connection From Present Predicted Cycle to Mean of Next Cycle

Robert L. Holland
Computer Sciences Corporation
Huntsville, Alabama

The following steps describe the procedure for developing the cubic connection from present predicted cycle to mean of next cycle:

1. Depending on when the present cycle was initialized, select the last inflection point of the estimated (present) cycle (t_i, F_i) near the 24-month point before the next cycle maximum or minimum. The 24-month limit was used to avoid a sudden rise or fall in the transition gap, since more than one inflection point exist in the prediction data. This should be done with at least a five-point numerical scheme for determining where the second derivative goes to zero^{B-1}. One such scheme is Sterling's in reference^{B-2}

$$\left(\frac{d^2 F}{dt^2} \right)_N = \frac{1}{(\Delta t)^2} \left[-\frac{F_{N+2}}{12} + \frac{4F_{N+1}}{3} - \frac{5F_N}{2} + \frac{4F_{N-1}}{3} - \frac{F_{N-2}}{12} \right]. \quad (B-1)$$

2. Select the maximum (or minimum) on the mean cycle (t_m, F_m). This is normally the input, but it can also be determined numerically.

3. The second derivative should be zero at (t_i, F_i), and the first derivative is zero at (t_m, F_m). A cubic curve is the lowest degree polynomial which can be determined that will go through the two points and have these properties at the two points.

4. The coefficients of the polynomial

$$F(t) = at^3 + bt^2 + ct + d \quad (B-2)$$

are to be determined from the following four linear equations and will have the required four properties:

$$6at_i + 2b = 0 \quad \{ \text{second derivative is zero at inflection point} \} \quad (B-3)$$

$$3at_m^2 + 2bt_m + c = 0 \quad \{ \text{first derivative is zero at maximum or minimum} \} \quad (B-4)$$

$$at_i^3 + bt_i^2 + ct_i + d = F_i \quad \{ \text{curve goes through the inflection point } (t_i, F_i), \} \quad (B-5)$$

$$at_m^3 + bt_m^2 + ct_m + d = F_m \quad \{ \text{curve goes through the maximum (or minimum) point } (t_m, F_m) \} \quad (B-6)$$

There are four linear equations in four coefficients which may be written in vector/matrix notation and solve for a, b, c, and d. Define vector F, matrix (M) and vector of coefficients to be determined c as follows:

$$\mathbf{F} \equiv \begin{pmatrix} 0 \\ 0 \\ F_I \\ F_m \end{pmatrix}, \quad \mathbf{c} \equiv \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} \quad \text{and} \quad (\mathbf{M}) \equiv \begin{pmatrix} 6t_I & 2 & 0 & 0 \\ 3t_m^2 & 2t_m & 1 & 0 \\ t_I^3 & t_I^2 & t_I & 1 \\ t_m^3 & t_m^2 & t_m & 1 \end{pmatrix}.$$

The corresponding vector/matrix equation for equations (B-3) through (B-6) above is

$$(\mathbf{M})\mathbf{c} = \mathbf{F} \quad (\text{B-7})$$

To solve for \mathbf{c} multiply the inverse $(\mathbf{M})^{-1}$ i.e.

$$(\mathbf{M})^{-1}(\mathbf{M})\mathbf{c} = (\mathbf{M})^{-1}\mathbf{F}$$

$$\text{then } \mathbf{c} = (\mathbf{M})^{-1}\mathbf{F}. \quad (\text{B-8})$$

Note that all the elements of $(\mathbf{M})^{-1}$ and \mathbf{F} are known data.
The solution for \mathbf{c} becomes

$$\mathbf{c} = \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} \frac{1}{3t_I}b \\ \frac{F_m - F_I}{-2t_m^2 + \frac{2}{3}t_m^3 - \frac{2}{3}t_I^2 + 2t_m t_I} \\ -\left(2t_m - \frac{t_m^2}{t_I}\right)b \\ F_I - \left(\frac{2}{3}t_I^2 - 2t_m t_I + t_m^2\right)b \end{pmatrix} \quad (\text{B-9})$$

References

- B-1. Waldmeier, M.: *The Sunspot Activity in the Years 1610-1960*. Zurich Schulthess and Company, Switzerland, 1961.
- B-2. Scarbrough, J. B. : *Numerical Mathematical Analysis*. Johns Hopkins Press, Baltimore, Maryland, 1962, pp. 133-134.

APPENDIX C

Estimates of Percentiles for the 13-Month Smoothed Solar Flux Data

Leonard W. Howell, Jr.
Electromagnetics and Aerospace Environments Branch
Systems Analysis and Integration Laboratory
George C. Marshall Space Flight Center
National Aeronautics and Space Administration

The intent of the MLLRT Model and the periodic memos was to provide NASA programs with an estimation of the most probable solar activity trends, based on statistical analysis of the historical record, with confidence bounds placed on that estimate. However, it has turned out that many of the most important applications required estimates for the high extreme values, and so the 97.7 percentile envelope was used in this application. For this reason a re-examination of the methodology for evaluating the percentile estimates associated with these bounding values was undertaken and some changes have been made. This appendix summarizes the results of this evaluation and the rationale for the revised approach.

Percentiles are the probability, expressed as a percentage, that a variable X will be less than a specified value x_p . To eliminate repetition on the factors of 100 in the mathematical expressions which follow, we denote this probability as a fraction, or "quantile", rather than a percentile. The only difference between percentile and quantile is that percentile refers to a percent of the set of data and quantile refers to a fraction of the set of data. For example, for a standard normal distribution, 1.96 is the 0.975 quantile and is the 97.5th percentile thus $p = \Pr\{X < x_p\}$ and it ranges from zero to one. The inverse function $Q(p) = x_p$ is also useful to yield the value in the range (F10.7 flux, in this case) associated with a given probability (p value). The concept of a quantile is illustrated in Figure C-1 for an arbitrary probability density function $f(x)$ and its cumulative distribution $F(x)$.

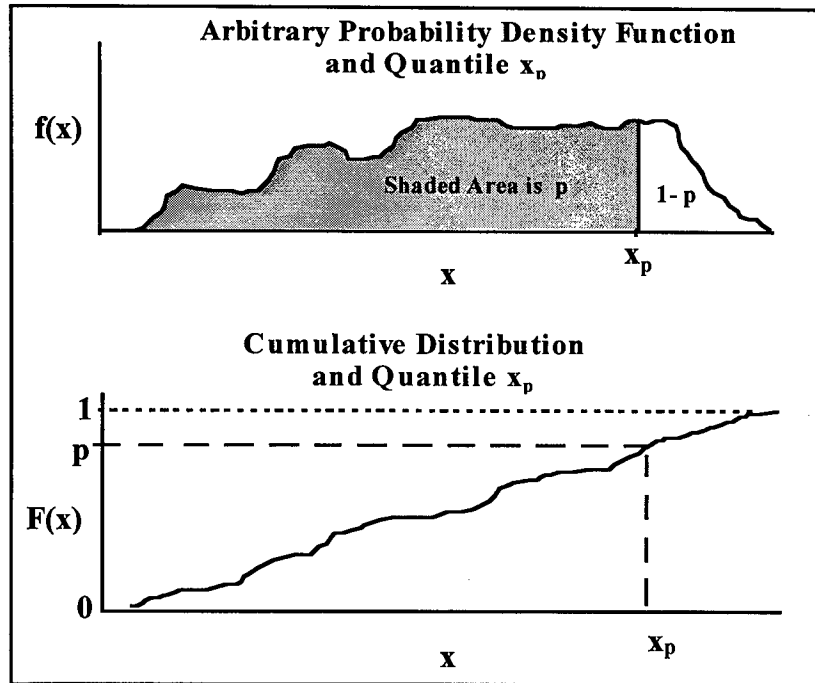


Figure C-1. Arbitrary Probability Density Function $f(x)$ and Its Cumulative Distribution $F(x)$.

Historically the MSFC solar activity memos have reported p values of 0.5, the median or “best estimate,” and bounding values of 0.023 and 0.977. These are easy to interpret for the engineering community because of their relationship to the “2-sigma” probabilities of the Normal distribution, even though there are strong indications that the Normal distribution is not applicable to this data set.

Calculation of percentile probabilities is a well defined process in cases where there is a large data set or where, for other reasons, the sample is known to follow a specified frequency distribution. The probability is the integral under the frequency distribution from the lower limit of the distribution, often $-\infty$, to x_p . For the problem at hand, however, where the data set consists of only 21 data points and the appropriate distribution is unknown, the probability can only be estimated. One must select one of the various possible methods. Direct calculation of a precise value which “truly” represents the situation is not possible.

For illustration we will use the 13-month smoothed Lagrangian flux data set illustrated in Figure C-2. The 13-Month Smoothed Lagrangian Solar Flux data is calculated from the monthly solar flux data that is first smoothed by the Zurich smoothed equation. The 13-month smoothed time series data is then distributed into groups of cycle data. The data then is interpolated by a Lagrangian method into 21 “peak-to-peak” solar cycles, with each cycle consisting of 132 points (months) as depicted in Figure C-2. It should be noted that the data is not really data in the usual statistical sense of “observations”, but it is actually the output from multiple layers of processing of the original observations. This fact is not relevant to the issues addressed in this appendix, but it is important to the physical interpretation of the solar model results. The processing tends to mask some of the actual variations and properties of the original data.

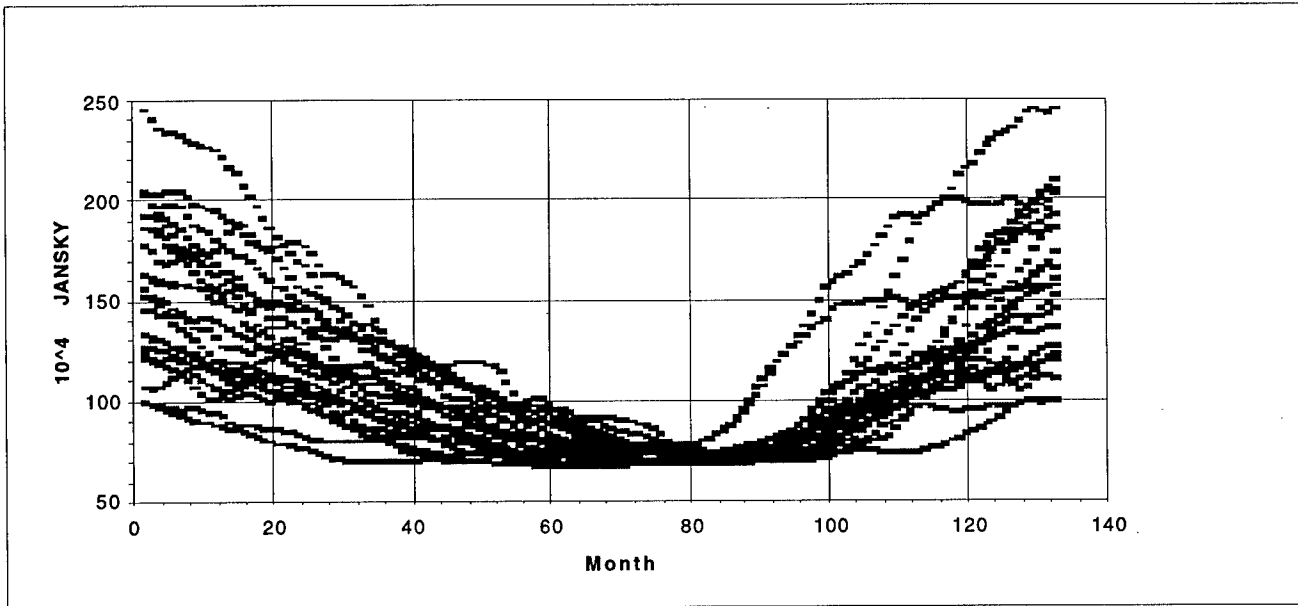


Figure C-2. 13-Month Smoothed Lagrangian Flux.

When observing this data, it is important to note that two cycles appear to run side-by-side and on the high side from month 80 to just under month 100, and then again briefly around months 115-118. Secondly, the mean and standard deviation trendlines can be constructed from the data and are shown in Figure C-3.

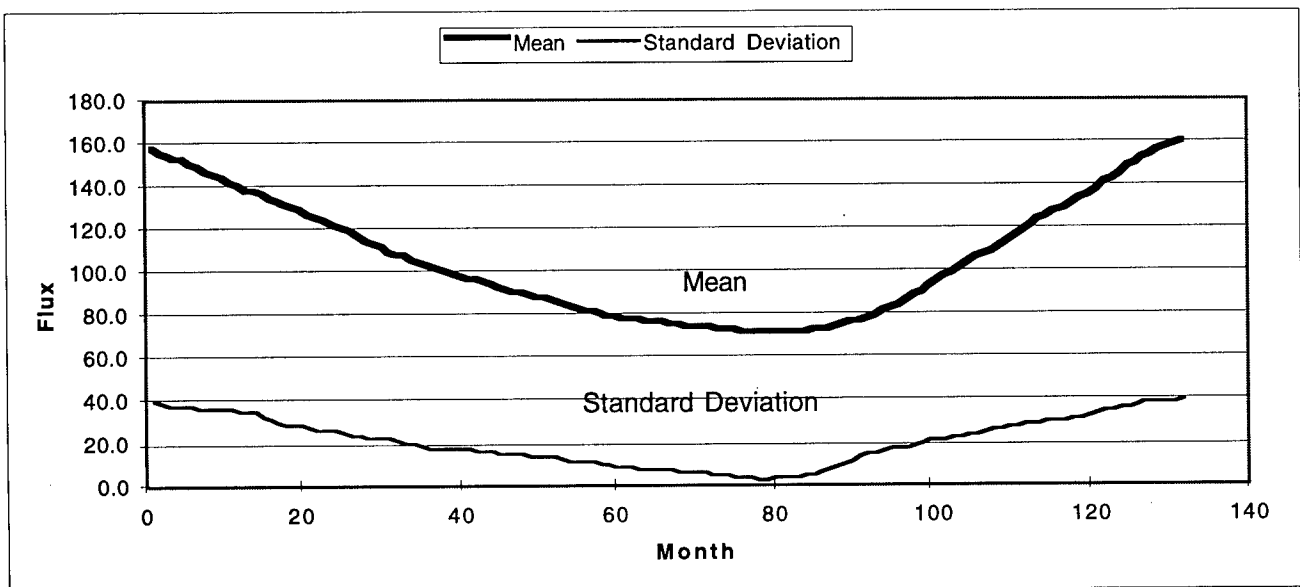


Figure C-3. Mean and Standard Deviation Trend Lines.

General

Two methods for estimating percentiles are developed and compared. Both methods presented make no attempt to analytically model an underlying distribution and thus represent strictly empirical approaches. The first method discussed makes use of frequency histograms to construct cumulative frequency polygons and is similar to the previous method. The second method presented is referred to as the Quantile method in the statistical literature and will be compared to the cumulative frequency polygon method.

Method 1. Cumulative Frequency Polygons

Graphical techniques can be easily implemented and they are very valuable in exploratory data analysis and in combination with formal statistical tests. In the former the objective is to discover particular features of the underlying distribution such as outliers, skewness or kurtosis (i.e., thickness of the tails of the distribution). The saying that a picture is worth a thousand words is especially appropriate for this type of analysis.

The frequency polygon is a many sided figure that is often used as an approximation of the underlying probability density function of a random variable and is constructed from a histogram of the data. Construction and use of the frequency polygon is well documented in the literature, and the interested reader is referred to Kohler [C-1], Downie and Heath[C-2], and Kendall and Stewart [C-3]. To draw the frequency polygon, the same set of coordinates is used as for the histogram, but this time the class mark, or midpoint of each class width, is identified (as the average of the two class limits) and a dot is positioned above it at a height equal to the relative frequency density. The dots are then connected by straight lines. To complete the polygon, straight lines are also drawn from the dots above the first and last class marks, respectively, to points on the horizontal axis a distance W below the first class midpoint and a distance W above the last class midpoint, where W is the class width.

This method is illustrated using the 21 data points defined by the first month of the 21 solar cycles. The data in rank order is given below:

$S = \{98.3, 99.9, 107.1, 120.5, 123.2, 125.4, 132.9, 133.2, 145, 152.1, 155, 162.2, 162.3, 177.4, 185.8, 186.4, 192, 202.2, 203.6, 203.7, 245.1\}$.

First, the sample mean is computed as 157.8 and the standard deviation as 39.7; these points are visible as the initial points (extreme left hand end) of the curves in Figure C-3. Next, the standardized data set is constructed as $z_i = (x_i - 157.8)/39.7$ which gives

$S_{\text{standardized}} = \{-1.498, -1.458, -1.276, -0.939, -0.871, -0.816, -0.627, -0.619, -0.322, -0.143, -0.07, 0.111, 0.114, 0.494, 0.706, 0.721, 0.862, 1.119, 1.154, 1.157, 2.2\}$

The frequency histogram and polygon are then constructed as illustrated in Figure C-4 for this standardized data set.

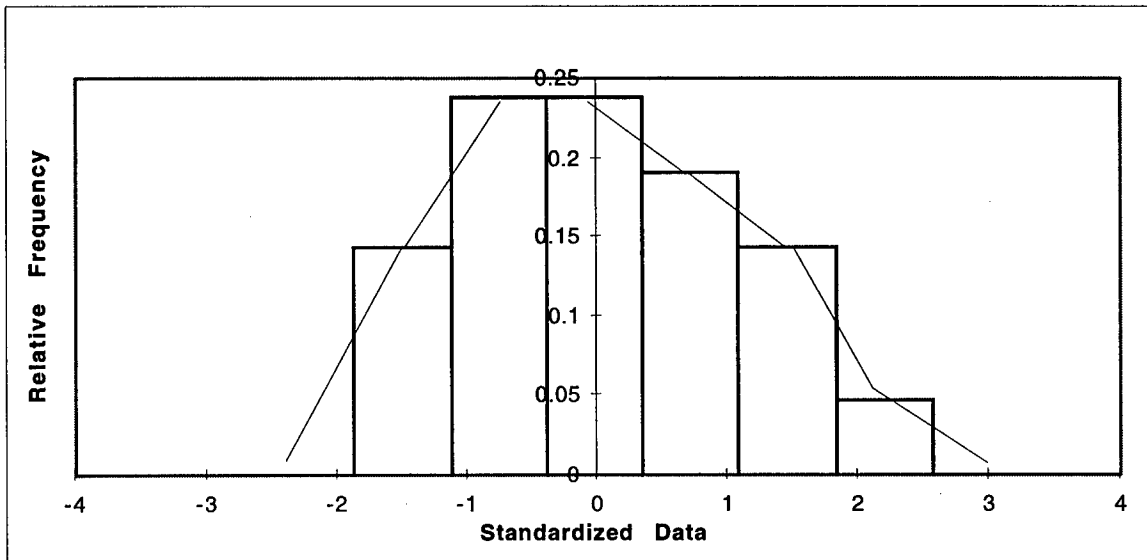


Figure C-4. Relative Frequency Histogram.

The cumulative frequency polygon or “less-than” ogive is then constructed directly from the polygon, as shown in Figure C-5. Quantiles of interest are then approximated using linear interpolation. More sophisticated methods may be applied by first smoothing the frequency polygon and then using higher order interpolation methods if desired.

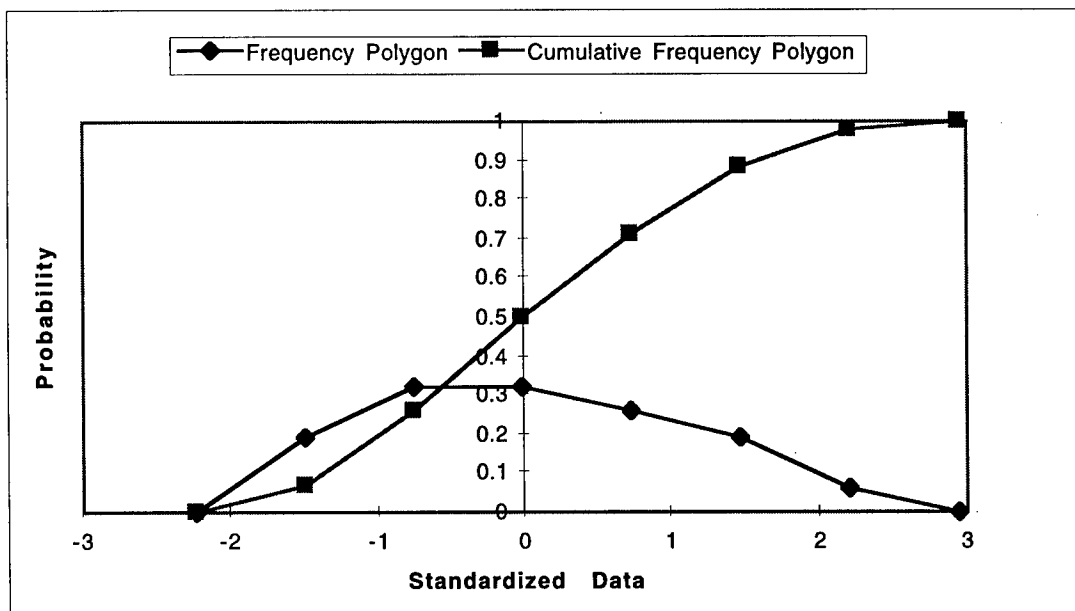


Figure C-5. Frequency Polygon.

Quantiles may be obtained from Figure C-5 using linear interpolation which are then transformed back to the original units of the data by multiplying by the standard deviation and then adding the

mean. For example, $Q(0.95)$ is $1.996 \times 39.7 + 157.8 = 237$, or, in terms of percentiles, 237 is the 95th percentile.

Discussion

When constructing the relative frequency polygon, one must first construct the relative frequency histogram. Constructing histograms usually requires a certain amount of personal judgment. For example, the number of classes to be used and the corresponding class width is a matter of personal choice. General guidelines recommend between five and twenty classes; a smaller number sacrifices too much detail, a larger one retains too much of it. Many statisticians use Sturges's rule to determine the number of classes k , where Sturges's rule gives $k = 1 + 3.3 \log_{10}(n)$, where n is the sample size. Another rule found in Panofsky and Brier [C-4] suggests $k = 5 \log_{10}(n)$. For the data set under consideration, $n = 21$ so the two rules give 5.5 and 6.6, and so $k = 6$ is a reasonable number of classes to begin with.

Next, the class width W must be determined. Assuming that the observations have been put in rank order so that $x_1 < x_2 < \dots < x_n$, then the range of the data is defined to be $R = x_n - x_1$, and hence the class width W must be greater than or equal to R/k in order to include all the data in the frequency count. However, if one simply sets $W = R/k$ then the endpoints of the data (x_1 and x_n) will correspond to the class limits of the first and last class, respectively, which is usually not appropriate. In fact, the general guidelines are that the class midpoints should roughly correspond to the mean of the data points that fall within each class. However, this in practice can only be achieved in some approximate sense so that it is very unlikely that two statisticians would ever arrive at the same histogram parameters for the same data set.

This is a very important consideration when implementing a general purpose algorithm for estimating quantiles and so we next explore the sensitivity of quantile estimation with respect to the class width, keeping the number of classes fixed at $k = 6$ for this data set. Two methods of class width selection are compared.

The first method selects $W = R/(k-1)$ which forces the endpoints of the data to be the midpoints of the first and last class. This method is applied across the 132 months, at each month constructing the associated cumulative frequency polygon and then estimating $Q(0.95)$ and $Q(0.976)$. A second method is chosen where the mean of the data (which is zero for the standardized data set) corresponds to one of the class midpoints and $W = (R+1)/k$ as recommended in Downie and Heath [C-2] which insures that x_1 and x_{21} do not correspond to the first and last class limits. Thus, with W defined, the number of classes set at $k = 6$, and one class midpoint defined, all class limits are automatically determined and the histogram and corresponding cumulative polygon can be constructed and the quantiles of interest estimated. $Q(0.95)$ and $Q(0.976)$ are compared for these two methods in Figures C-6 and C-7; the monthly mean curve is included in each figure as a reference curve. Notationally, the graph labeled Polygon2 refers to the $W = R/(k-1)$ method. Now, in both of these figures we note that while the two graphs are fairly close to each other, they nevertheless differ in a number of places. Most notably, however, is that each graph has some unusual jagged sections in one region but does not necessarily have a jagged section in the graph produced by the other method in the same region.

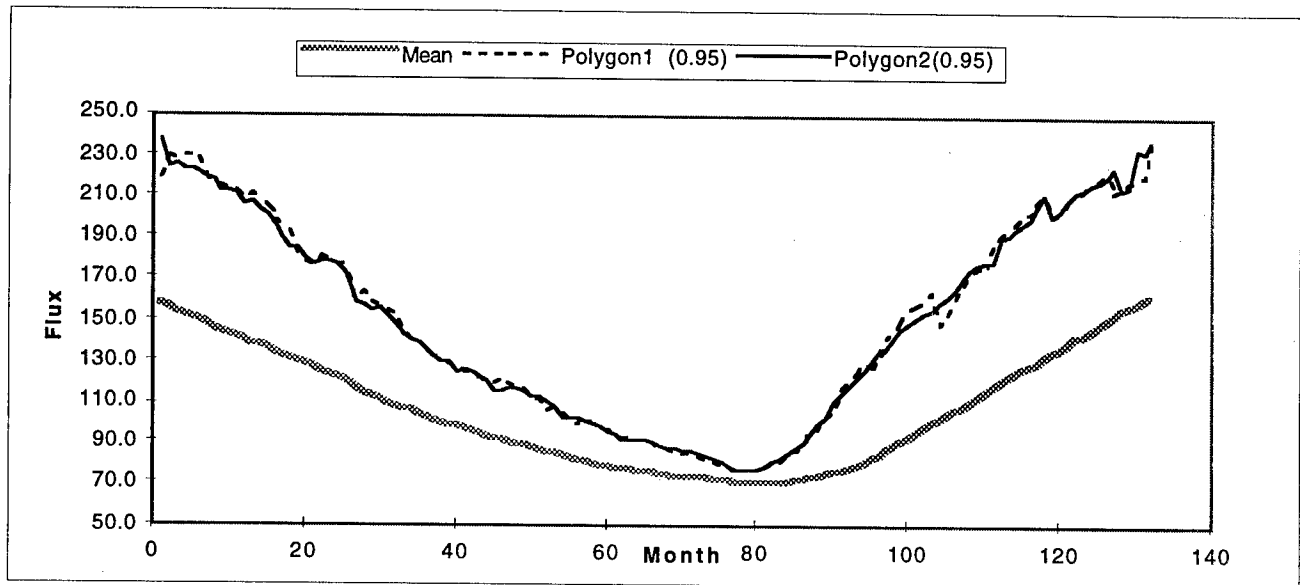


Figure C-6. Comparison of Frequency Polygon Methods for $Q(0.95)$.

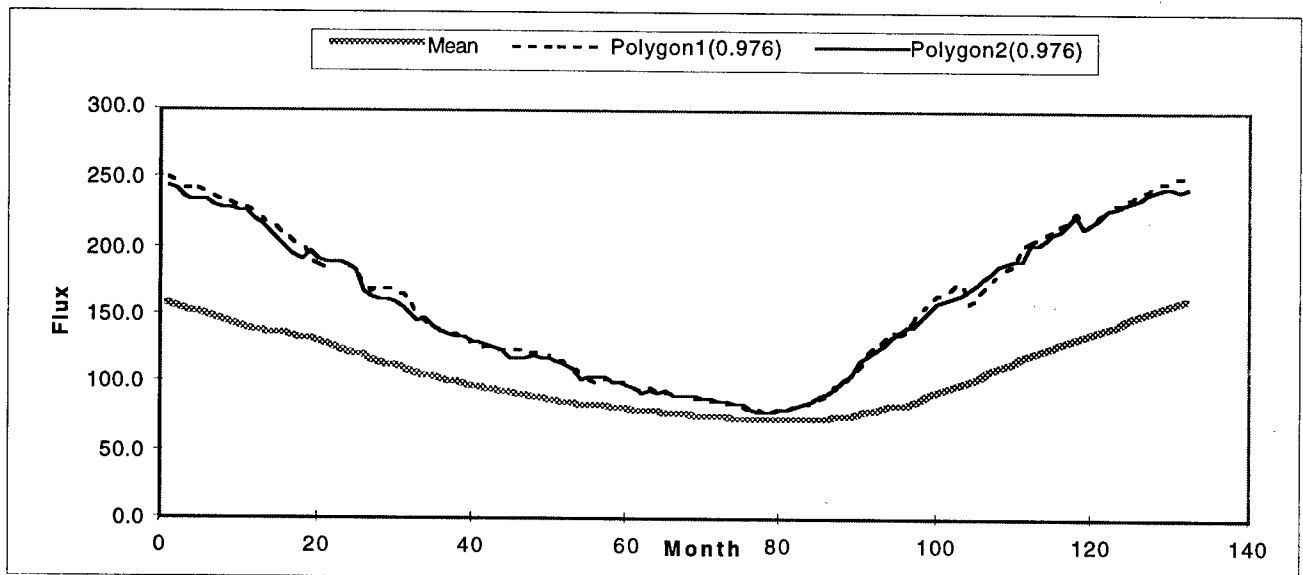


Figure C-7. Comparison of Frequency Polygon Methods for $Q(0.976)$

This clearly points out how sensitive quantile estimation is to the choice of the class width W . Thus, given that the selection of W (and even k , the number of classes) is quite subjective and variations in W produce different and even jagged results in Q , it is strongly recommended that quantiles NOT be estimated using frequency polygons. However, this does not detract from their use for other purposes, such as their graphical value.

Method 2: Quantile Plots

Chambers, et. al. [C-5], give a succinct discussion of using quantile plots and the quantile function in their book *Graphical Methods for Data Analysis*. As with Method 1, these methods will be illustrated using the 21 data points defined by the first month of the 21 solar cycles. The actual data presented in rank order is given below:

$S = \{98.3, 99.9, 107.1, 120.5, 123.2, 125.4, 132.9, 133.2, 145, 152.1, 155, 162.2, 162.3, 177.4, 185.8, 186.4, 192, 202.2, 203.6, 203.7, 245.1\}$,

and recall that the sample mean is 157.8 and the standard deviation is 39.7.

The authors in [C-5] next construct an operational definition of quantile and give supporting rationale for its functional form. Starting with a set of rank ordered data given as x_i for $i=1$ to n , with $x_1 < x_2 < \dots < x_n$, let p represent any fraction between 0 and 1. Then the quantile $Q(p)$ corresponding to the fraction p is defined to be x_i whenever p is one of the fractions $p_i = (i-0.5)/n$, for $i=1$ to n .

Thus, for the 21 observations in S , the following table constructed, where $p_i = (i-0.5)/21$ for $i=1$ to 21.

p_i	0.024	0.071	0.119	0.167	0.214	0.262	0.310	0.357	0.405	0.452	0.500
$Q(p_i)$	98.3	99.9	107.1	120.5	123.2	125.4	132.9	133.2	145.0	152.1	155.0
p_i	0.500	0.548	0.595	0.643	0.691	0.738	0.786	0.833	0.881	0.929	0.976
$Q(p_i)$	155.0	162.2	162.3	177.4	185.8	186.4	192.0	202.2	203.6	203.7	245.1

This may then be presented as a quantile plot as depicted in Figure C-8.

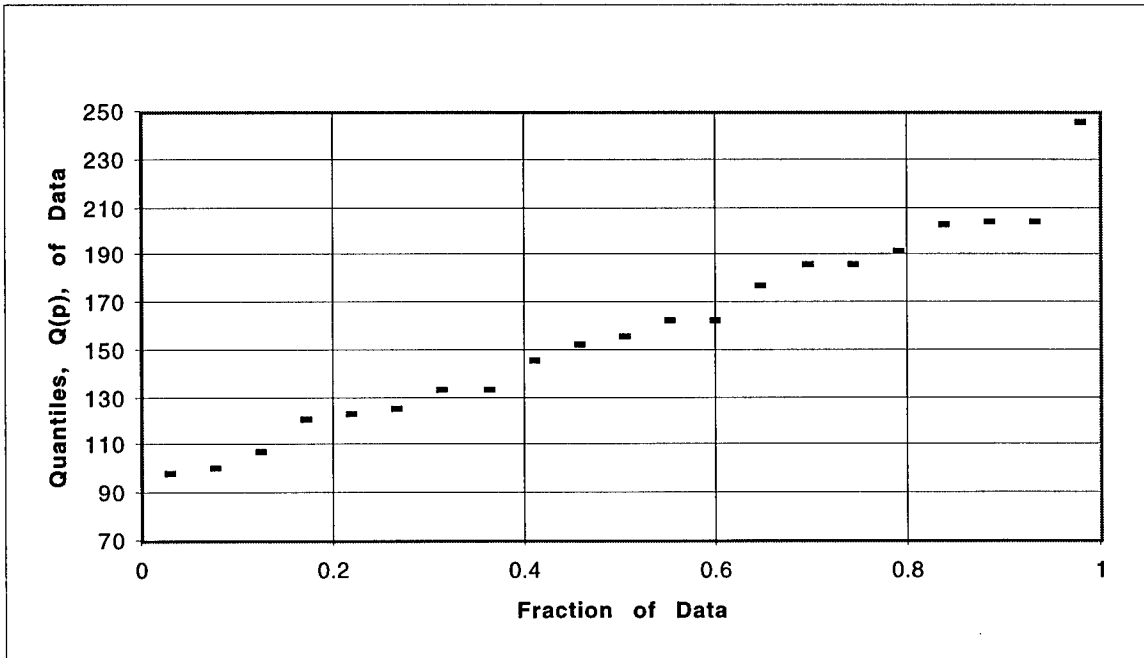


Figure C-8. Quantile Plot.

So far, the quantile function is defined only for certain discrete values of p , namely p_i . If necessary, the definition of $Q(p)$ is extended within the range of the data by linear interpolation. This means connecting consecutive points with straight line segments, as show in the Figure C-9 below. Symbolically, if p is a fraction f of the way from p_i to p_{i+1} , then $Q(p)$ is defined as

$$Q(p) = (1-f)Q(p_i) + f Q(p_{i+1})$$

This cannot be used to define $Q(p)$ outside the range of the data, where p is smaller than $0.5/n$ or larger than $(1-.5/n)$. As the authors in [C-5] state:

“Extrapolation is a tricky business; if we must extrapolate we will play safe and define $Q(p)=x_1$ for $p < p_1$ and $Q(p)=x_n$ for $p > p_n$, which produces the short horizontal segments at the beginning and end of the Figure C-9.”

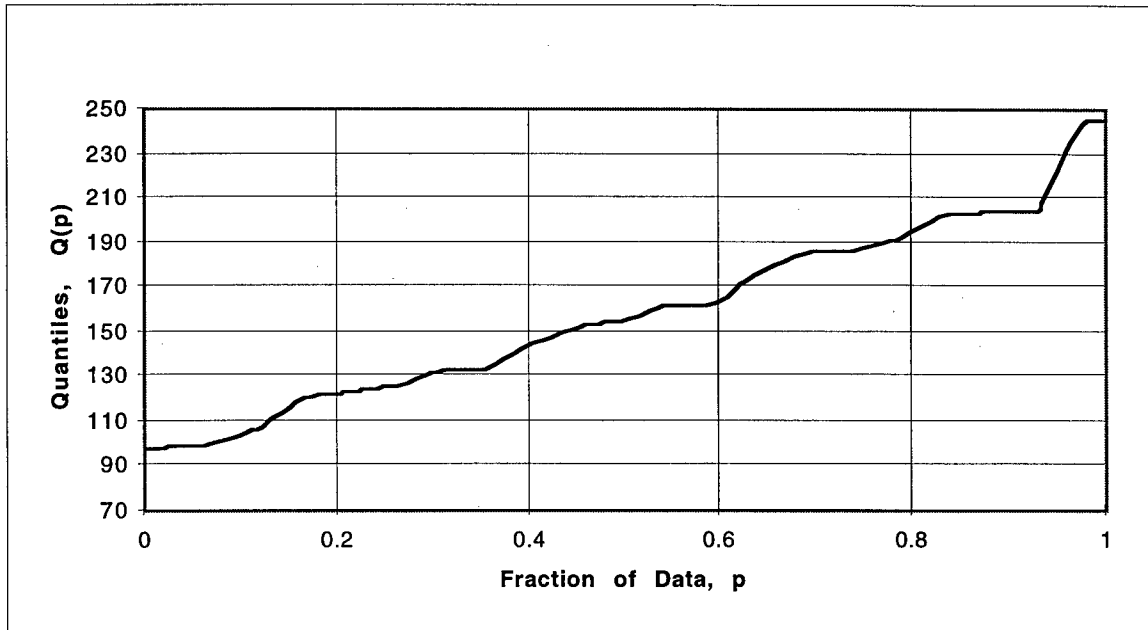


Figure C-9. Interpolated Quantiles.

In practical terms, quantiles smaller than 0.0238 (percentiles smaller than 2.38%) or larger than 0.976 (percentiles larger than 97.6%) are beyond the resolution of the sample size under consideration ($n=21$) and cannot be derived from the data without further assumptions. Thus, for this data set, $x_1=98.3$ is the 0.0238 quantile and $x_{21}=245.1$ is the 0.976 quantile. This result can be generalized to say that for any data set consisting of 21 observations, the minimum value x_1 is the 0.0238 quantile and the maximum value x_{21} is the 0.976 quantile.

In order to go beyond these limits, one approach we considered was to learn if the theoretical endpoints of the distribution of data, corresponding to $Q(0)$ and $Q(1)$, were known or could be estimated using physical laws. Linear interpolation could then be used to approximate the very small or very large quantiles of interest. Unfortunately this approach did not produce usable bounds.

The quantile method is applied to the 21 solar cycles at each of the 132 months to give $Q(0.95)$ and $Q(0.976)$, i.e., the 95th and 97.6th percentiles. These results are depicted in Figure C-10.

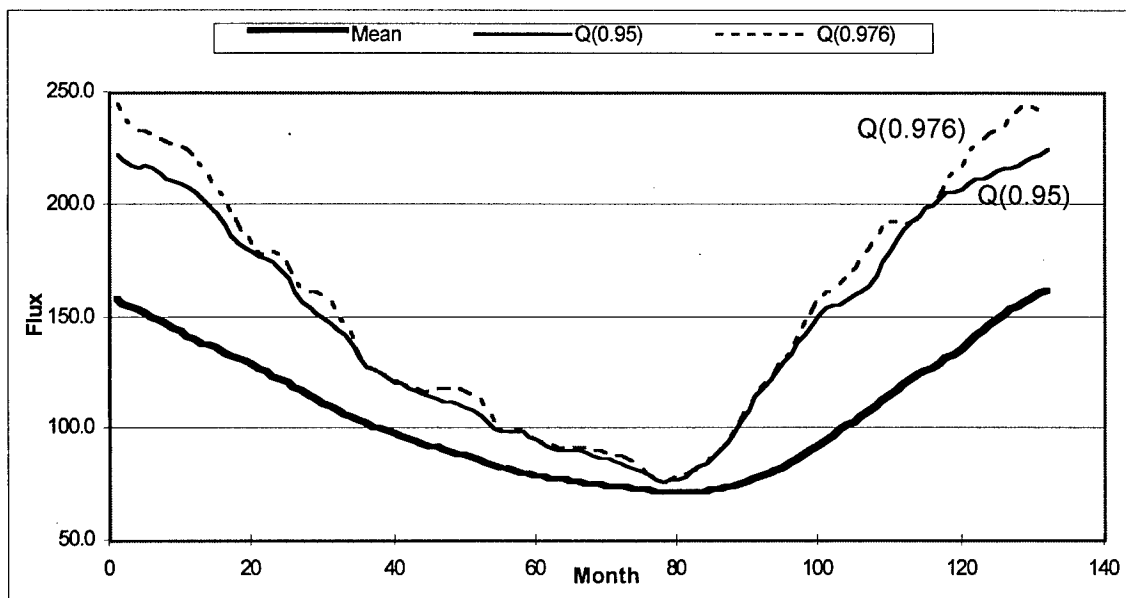


Figure C-10. Quantile Method

Note that $Q(0.95)$ and $Q(0.976)$ are roughly identical in the months 80-95 and months 115-118 where we previously noted that two cycles appear to run side-by-side and on the high side. This is because $Q(0.95)$ is computed by interpolating between x_{20} and x_{21} which are roughly equal to each other in these regions. Furthermore, $Q(0.95)$ yields a somewhat smoother graph than $Q(0.976)$ as one would expect because $Q(0.95)$ is determined by linear interpolation between x_{20} and x_{21} (which is a form of averaging and hence smoothing), while $Q(0.976)$ is always the maximum of the data (x_{21}) and is therefore more sensitive.

The Quantile Method and Program Risk

As stated previously, Chambers, et al.[C-5], define the quantile $Q(p)$ corresponding to the fraction p to be x_i whenever p is one of the fractions $p_i=(i-0.5)/n$, for $i=1$ to n . This definition of Q is widely used and the authors support their choice of the functional form of Q as follows:

“Why do we take p_i to be $p_i=(i-0.5)/n$ and not, say i/n ... or several other reasonable choices? We will mention only that when we separate the ordered observations into two groups by splitting exactly on an observation, the use of the function $(i-0.5)/n$ means that the observation is counted as being half in the lower group and half in the upper group.”

This is a very interesting, if not compelling, argument in favor of this functional form, and would usually be used without further consideration in most applications, especially in regions away from the endpoints of the data. However, it is important to look at the application of these results to NASA programs and the inherent program risks associated with their use. For example, as discussed in the preceding paragraphs, this functional form for Q suggests that if we select the largest observation from a sample of size 21, then we define it to be the 0.976 quantile, or in programmatic terms, we are stating that the risk is only 2.4% and this result is based on a

sample of size 21, without any probability distribution assumptions! Because significant program design and planning decisions may potentially be based on these numbers, a deeper investigation of their properties is appropriate.

First, note that the quantile function described above is associated with a special case of the so-called empirical distribution function (EDF), which is a step function generally defined by

$$F_n(x_i) = \frac{i - c}{n - 2c + 1} \quad \text{for } 0 \leq c \leq 1 \quad (\text{C-1})$$

for the ordered observations $x_1 \leq x_2 \leq \dots \leq x_n$. Figure C-11 illustrates the inverse relationship between the quantile function, Q , and the empirical distribution function, F_n , where p_i is defined to be $p_i = (i - c)/(n - 2c + 1)$.

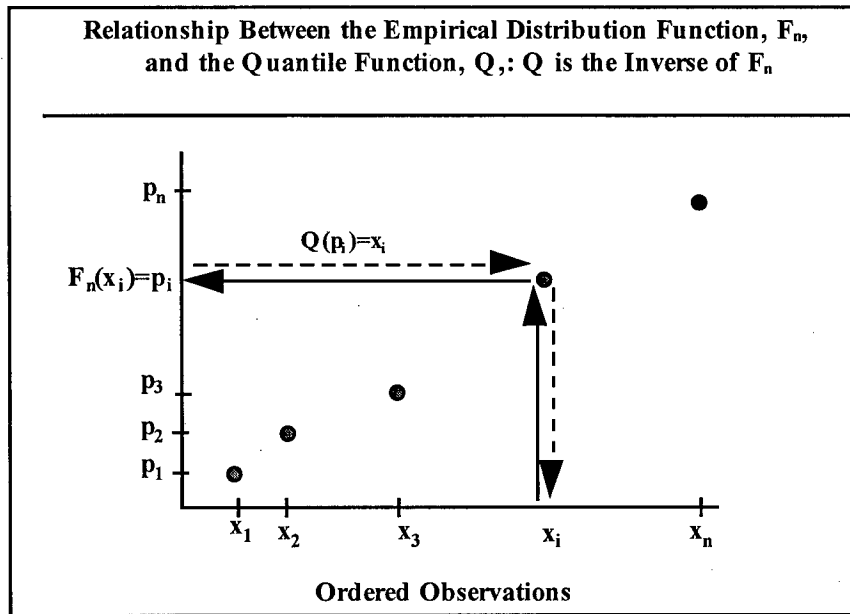


Figure C-11. Relationship Between the Empirical Distribution Function, F_n , and the Quantile Function, Q .

The objective at this point is to select a value of c that makes the most sense for our application. However, it must be understood that whatever methods employed to select c must be quite general in that they are distribution free, i.e., there will be no probability distribution assumptions made. This is necessary because the underlying probability distribution of the solar flux under consideration is obviously not known; otherwise the quantiles would also be known exactly and there would be no reason for analysis. With this in mind, a branch of statistics referred to as "order statistics" will be used.

For a sample of n values x_1, x_2, \dots, x_n with a probability density function $f(x)$ and cumulative distribution $F(x)$, and where the observations have been arranged in rank order so that $x_1 \leq x_2 \dots \leq x_n$, then the probability density function of the i^{th} order statistic x_i is given by the following equation, as derived in Kendall and Stewart [C-3],

$$g(x_i) = \frac{n!}{(i-1)!(n-i)!} \{F(x_i)\}^{i-1} \{1-F(x_i)\}^{n-i} f(x_i) \quad (\text{C-2})$$

Thus, for any distribution with density function $f(x)$ and cumulative distribution $F(x)$, if a random sample of size n is selected from this distribution and the observations are arranged in increasing order, the above equation defines the probability density function of the i^{th} order statistic.

For example, the probability density function of the minimum or smallest observation is obtained by setting $i=1$ in Equation C-2, giving $g(x_1) = n\{1-F(x_1)\}^{n-1} f(x_1)$. Similarly, the probability distribution of the maximum or largest observation is obtained by setting $i=n$ in Equation C-2, giving $g(x_n) = n\{F(x_n)\}^{n-1} f(x_n)$. Now, in the solar flux study, nature (and processing of the data) provides a random sample of size n ($n=21$ at present) for each of 132 months, and then they are arranged in rank order to construct the quantile plot. Since a linear combination of these ranked observations is to be used to provide a quantile associated with a particular p -value, the distribution of the p -values that are determined by a linear combination of the order statistics shall be investigated.

That is, if a linear combination of the ranked observations is used to estimate a particular quantile x_p of interest, the question "what is the distribution of p -values provided by this linear combination of the ordered observations?" must be answered. For example the maximum observation x_{21} is begin considered to estimate a p -value in the neighborhood of 0.97. Therefore, the following question is to be answered: what is the distribution of p -values obtained when x_{21} is used to estimate the percentile of interest?

Imagine that as nature generates or picks the 21 observations from the unknown distribution, the maximum value x_{21} is selected and the question is "what is the p -value associated with this particular x_{21} ?" The answer is $F(x_{21})$, i.e., the area under the density function $f(x)$ curve and to the left of x_{21} as illustrated in Figure 1. Repeating this sampling process of selecting another set of 21 observations and picking the maximum value x_{21} of each of these sets of 21 observations, one would like to know the distribution of the p -value associated with this set of x_{21} 's. That is, what is the probability distribution of $p_{21}=F(x_{21})$ and more generally, the distribution of $p_i=F(x_i)$ for each of the order statistics?

First, note that the p 's and the x 's are in the same order since p is a non-decreasing function of x , and furthermore $0 \leq p_1 \leq p_2 \dots \leq p_n \leq 1$ since F is bounded by 0 and 1. As discussed above, the objective at this point is to discover the probability distribution of the order statistics p_i . To help answer this question, a famous result in probability theory, often called the Probability

Integral Transformation $p=F(x)$, is used. This theorem proves that the distribution of p is the standard uniform distribution, independent of the distribution $F(x)$. That is, $f(p)=1$ for $0 \leq p \leq 1$ and $F(p)=p$. In fact, this is the underlying theory behind random number generation on digital computers, which is often use in Monte Carlo simulations. The computer generates a standard uniform random number u between 0 and 1 and then solves the equation $x=F^{-1}(u)$ to generate random numbers from any specified distribution $F(x)$.

Inserting this result in the general formula for the distribution of the i^{th} order statistic given in Equation C-2, the probability density function of p_i is obtained and is seen to be Beta distribution:

$$f(p_i) = \frac{n!}{(i-1)!(n-i)!} p_i^{i-1} (1-p_i)^{n-i}, \quad 0 \leq p_i \leq 1, \quad (\text{C-3})$$

with mean and variance given by $\mu = \frac{i}{n+1}$ and $\sigma^2 = \frac{i(n-i+1)}{(n+1)^2(n+2)}$.

This is a very important result because if c is set to 0, i.e., $c=0$ in Equation (C-1), then

$$F_n(x_i) = \frac{i}{n+1} \quad \text{for } i = 1, 2, \dots, n \quad (\text{C-4})$$

which is identical to the mean of the distribution of p_i and explains why $c=0$ is called the “mean” plotting position. For example, suppose for the solar sample of size 21 the largest observed value x_{21} is used as the quantile. The question is, what is p_{21} ? By the first definition of Q it would be $p_{21}=Q(x_{21})=(21-.5)/21=0.976$. However, if Q is defined as $Q(x_i)=i/(n+1)$, then the definition of a quantile is forced to coincide with the mean of the distribution of p_i , so that $p_{21}=21/22=0.954$. To see which definition makes the most sense for the present application, the distribution of p_{21} may be analyzed by setting $i=n=21$ in the Beta distribution above, getting

$$f(p_{21}) = 21 p_{21}^{20}, \quad 0 \leq p \leq 1 \quad (\text{C-5})$$

As noted above, the mean is given by $21/22=0.954$, and the median is given by $\sqrt[2]{\frac{1}{2}}=0.968$. Furthermore, a 90% tolerance interval for p_{21} is determined in a similar manner and is given by **{0.867, 0.998}** with a midpoint of **0.932**. In comparing the above two definitions for Q , the choice $(i-0.5)/n$ gives a result closer to the median of the distribution, but the choice of $i/(n+1)$ yields the more reasonable result, in that it exactly matches the mean of the distribution and is also more central to the midpoint of the interval. A similar argument by symmetry holds for the minimum order statistic x_1 and its associated p_1 .

For quantiles not corresponding to x_1 and x_{21} , linear interpolation should be used as suggested by Chambers, et. al.,[C-5] between the appropriate values, and the quantile function will be defined

as $Q(x_i)=i/(n+1)$. Furthermore, the Beta distribution can then be used to give the exact distribution of the interpolated or "smoothed" p-values. As we've seen above, the means will always match.

The interested reader may find a computer simulation helpful in visualizing these results. A general procedure for conducting a computer simulation is outlined below and an example is provided:

Step 1: Generate 21 random numbers from any specified probability distribution. For example, generate 21 numbers from a standard normal distribution, where the random number generator used in this computer program generates a single standard normal distribution by summing 12 uniformly distributed random numbers and subtracting 6. This insures that the p-values obtained in Step 2 are not the same uniform random numbers used to generate the random sample in the first place.

Step 2: Order the 21 random numbers in ascending order. Then for each random number, compute its p-value, i.e., the area under the normal distribution and to the left of it. Store these results as Run 1 as in Table C-1.

Step 3: Repeat steps 1 and 2. This is done 5000 times in this example.

Step 4: Analyze the results. Relative frequency histograms of the p_1 , p_{11} , and p_{21} data in Table C-1 are constructed and compared to the Beta distribution in Equation C-3. These results are illustrated in Figures C-12, C-13, and C-14, where the Beta distribution is the smoothed curve shown with the histogram, showing excellent agreement between simulation results and the theoretical model. Use Table C-2 to compare between the means and variances of the simulated results for p_1 , p_{11} , and p_{21} and corresponding means and variances of the Beta distribution given in Equation C-3.

Run

Table C-1. Simulated Data and Sample Means and Variances

1	x1	x2																			x21
	-1.63	-1.42	-1.33	-0.70	-0.65	-0.63	-0.40	-0.37	-0.30	-0.22	0.00	0.10	0.20	0.23	0.24	0.41	0.44	0.80	0.86	1.26	2.06
	p1	p2																			p21
	0.052	0.078	0.092	0.243	0.257	0.263	0.343	0.354	0.384	0.414	0.499	0.541	0.579	0.591	0.594	0.659	0.670	0.788	0.805	0.895	0.980
2	x1	x2																			x21
	-1.68	-1.54	-1.41	-1.12	-1.08	-1.06	-0.61	-0.43	-0.34	-0.23	-0.20	-0.20	0.18	0.35	0.82	0.88	0.89	1.17	1.28	1.31	1.91
	p1	p2																			p21
	0.046	0.062	0.079	0.131	0.141	0.144	0.272	0.335	0.367	0.408	0.420	0.423	0.573	0.636	0.794	0.810	0.813	0.880	0.899	0.905	0.972
.																					
.																					
.																					
5000	x1	x2																			x21
	-2.21	-1.56	-1.47	-1.24	-0.71	-0.54	-0.34	0.04	0.06	0.18	0.40	0.42	0.54	0.55	0.64	0.76	1.08	1.12	1.31	1.34	2.07
	p1	p2																			p21
5000	0.014	0.060	0.070	0.107	0.237	0.294	0.369	0.518	0.526	0.572	0.656	0.661	0.705	0.710	0.738	0.778	0.860	0.870	0.905	0.909	0.981
sample mean	0.045	0.090	0.135	0.180	0.225	0.271	0.317	0.362	0.407	0.453	0.499	0.545	0.591	0.637	0.683	0.729	0.773	0.819	0.863	0.909	0.955
sample variance	0.002	0.004	0.005	0.007	0.008	0.009	0.010	0.011	0.011	0.011	0.011	0.011	0.011	0.010	0.010	0.009	0.007	0.006	0.005	0.003	0.002

Table C-2. Means and Variances of the Beta Distribution

Theoretical Mean and Variance of Beta Distribution (see Equation C-3)																					
i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
mean	0.045	0.091	0.136	0.182	0.227	0.273	0.318	0.364	0.409	0.455	0.500	0.545	0.591	0.636	0.682	0.727	0.773	0.818	0.864	0.909	0.955
variance	0.002	0.003	0.004	0.006	0.007	0.008	0.008	0.009	0.009	0.009	0.010	0.009	0.009	0.009	0.008	0.008	0.007	0.006	0.004	0.003	0.002

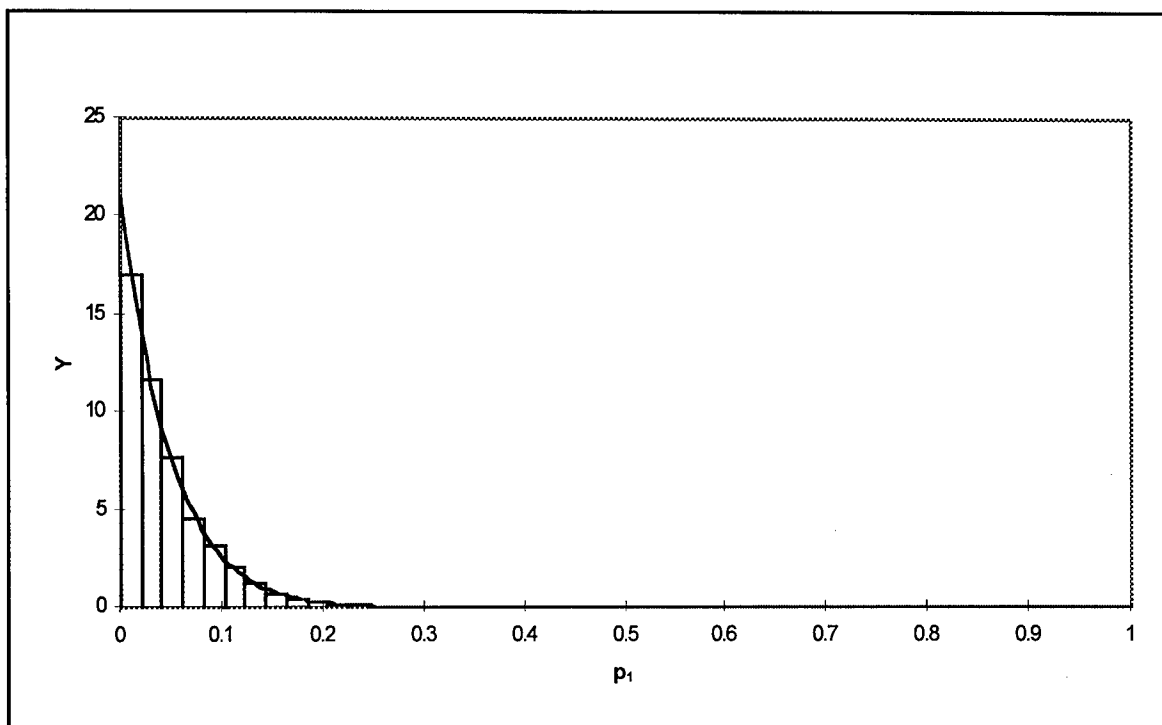


Figure C-12. Relative Frequency Histogram and Probability Density Function of p_1 .

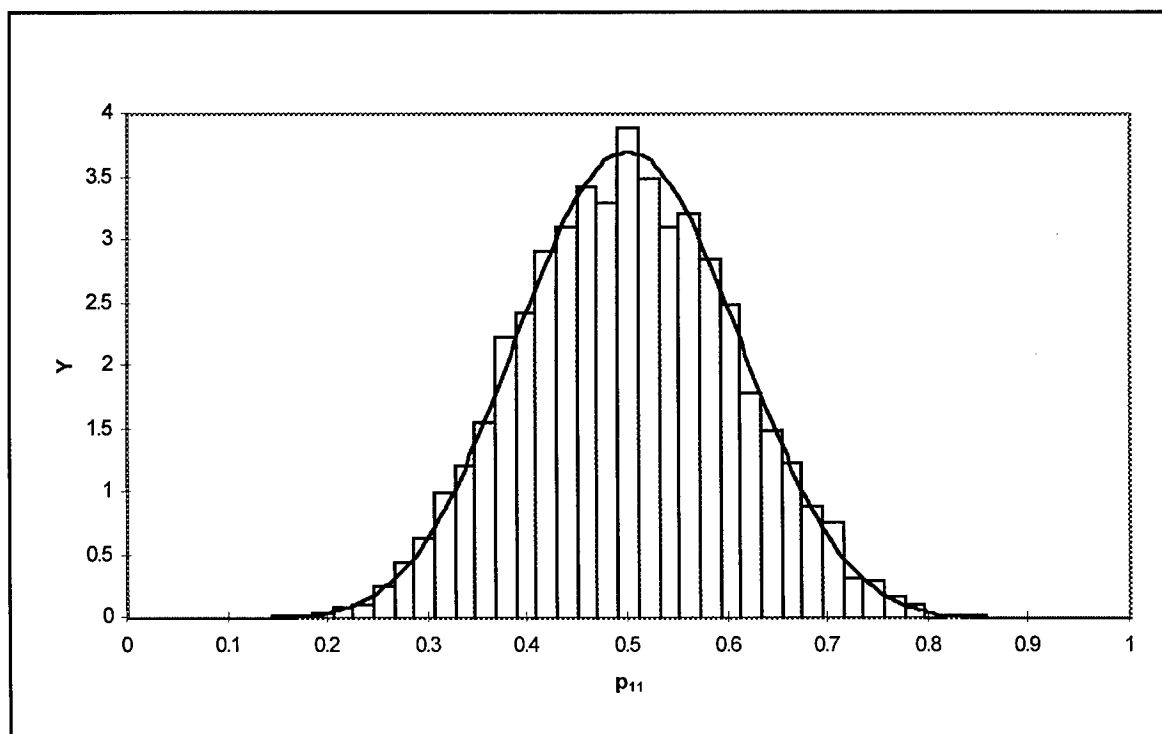


Figure C-13. Relative Frequency Histogram and Probability Density Function for p_{11} .

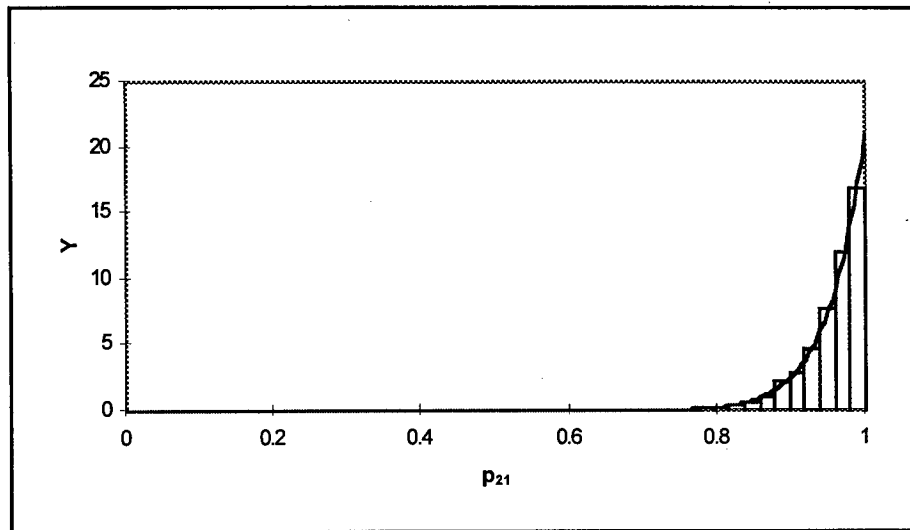


Figure C-14 Relative Frequency Histogram and Probability Density Function for p_{21} .

Continuing with a discussion of program risks, it is standard practice to talk about the Mean Time Between Failures (or the occurrence of some disastrous event), and hence using the mean $i/(n+1)$ equates correctly to this terminology. A similar argument in favor of $i/(n+1)$ versus $(i-0.5)/n$ by E. J. Gumbel [C-6] states the following about the method of using $(i-0.5)/n$ in a discussion of return periods:

the method “... claims that an event which has already happened once in n years will occur, in the mean, in $2n$ years. If the extreme observation has economic consequences, as in the case of floods, the danger factor is heavily under-estimated. The compromise is misleading where the plotting problem is of most interest.”

He continues on with his discussion of return periods of extreme events and concludes that $i/(n+1)$ is the most appropriate to use for these applications.

In conclusion, it is recommended that the following methods be used when using the quantile method for the present application:

- (1) Use the approach outlined by Chambers [C-5], but set $Q(p_i)=x_i$ when $p_i=i/(n+1)$ instead of $(i-0.5)/n$
- (2) use the linear interpolation method when necessary as previously discussed
- (3) recognize that percentiles smaller than 4.5% or larger than 95.4% are beyond the resolution of the sample size under consideration ($n=21$) and cannot be derived from the data without further assumptions.
- (4) use $Q(0.95)$ instead of $Q(0.954)$ because it is more common and will be smoother as a result of interpolation, as previously discussed.

Using these methods for the data set under consideration, $Q(0.95)=240.96$. Figure C-15 shows $Q(0.95)$ with these recommendations implemented. The mean is provided as a reference curve.

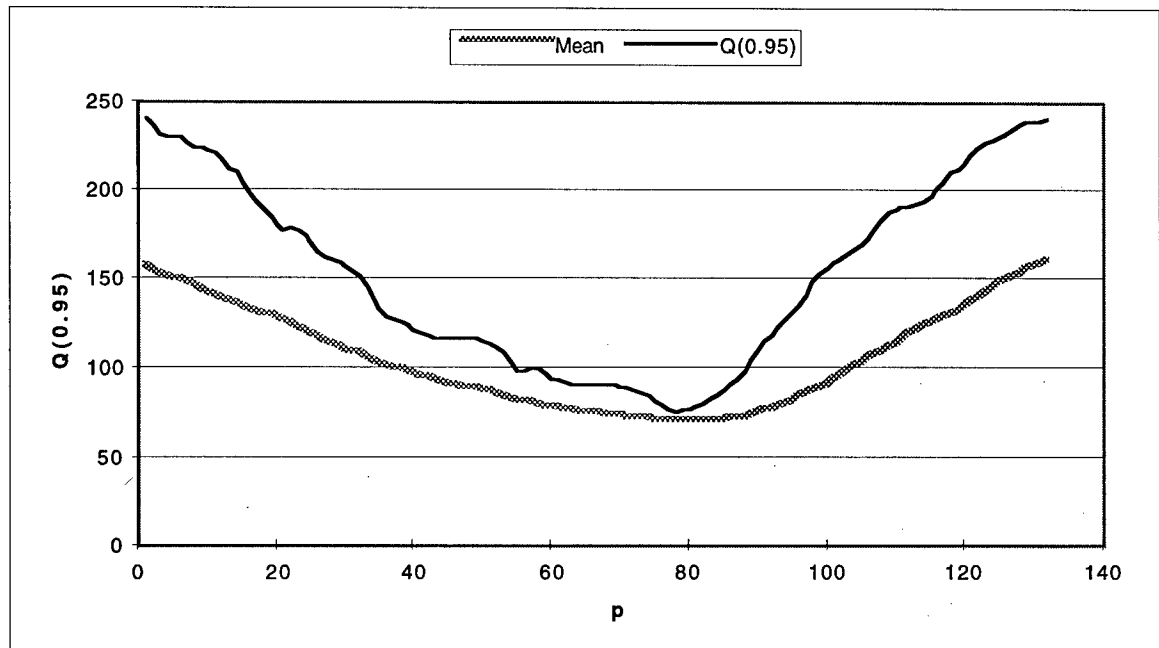


Figure C-15. Quantile Method: Q(0.95)

Recommendations

Two methods of estimating percentiles are investigated in this report. The quantile method and the relative frequency polygon method make no attempt to analytically model an underlying distribution and thus represent strictly empirical approaches. Of these two methods, the relative frequency polygon method is not recommended for use in estimating percentiles because of the sensitivity of $Q(p)$ to the class width and resulting jagged edges in Q .

Therefore, it is recommended that the Quantile method be used to estimate percentiles as long as they are within the resolution of the sample size. As stated earlier, this means that percentiles smaller than 4.5% and larger than 95.4% are beyond the resolution of the sample size under consideration ($n=21$) and cannot be derived from the data. Furthermore, we learned that $Q(0.95)$ yields a somewhat smoother graph because of the linear interpolation between x_{20} and x_{21} while $Q(0.954)$ is always the maximum of the data (x_{21}) and is therefore more sensitive. Therefore, it is recommended that the 95th percentile be estimated in the present application when using the quantile method.

References

- C-1. Kohler, Heinz, *Statistics for Business and Economics*, Scott, Foresman and Company, Glenview, Illinois, 1988, pp. 49-89.
- C-2. Downie, N. M., and Heath, R. W., *Basic Statistical Methods*, Harper and Brothers, New York, 1959, pp. 16-28.
- C-3. Kendall, Maurice G, and Stuart, Allan, *The Advanced Theory of Statistics*, Hafner Publishing Company, New York, 1963,
- C-4. Panofsky, Hans A., and Brier, Glenn W., *Some Applications of Statistics to Meteorology*, The Pennsylvania State University, University Park, PA, 1958, pp.2-13.
- C-5. Chambers, John M., et. al., *Graphical Methods for Data Analysis*, Wadsworth International Group and Duxbury Press, 1983, pp. 1-100.
- C-6. Gumbel, E. J., *Statistics of Extremes*, Columbia University Press, New York, 1960, pp. 4-34.

APPENDIX D

Modified McNish-Lincoln and Quantile Subroutine Examples

SUBROUTINE LINCMC (R,N,NU2)

```

      INTEGER*2 P,Q,IDG,K,I,J,M,MUPRIM,MU2P1,MU2,MU1,MU,NU1,NU2,I2,NU,N

      REAL*8 UPPER_PERCENTILE,LOWER_PERCENTILE,R(132,25),RMEAN(264),
      1RPRED(264),ERRUP(264),ERRDN(264),DR(132,25),A(20,20),AINV(20,20),
      2APRIME(100),B(132,25),C(132,25),JDUM(20),DETA(2),SIGM(132),SIGMA

      COMMON /BLK1/ I2
      COMMON /IDG/ IDG
      COMMON /ANS/ RMEAN,RPRED,ERRUP,ERRDN,SIGM
      COMMON /COUNT/ NU,NU1,MU,MU1,MU2,MU2P1,M
      COMMON /PRCNT/ UPPER_PERCENTILE,LOWER_PERCENTILE

      EQUIVALENCE (A(1,1),AINV(1,1),APRIME(1))

      MU1=MU2-MU+1
      NU=NU2-NU1+1
      MUPRIM=M-MU2

      IF (IDG.EQ.1.OR.IDG.EQ.2) THEN
        WRITE (44,400) 'YOU ARE IN LINCMC SUBROUTINE'
        WRITE (44,410) 'M= ',M,'N= ',N,'NU1= ',NU1
        WRITE (44,410) 'NU2= ',NU2,'MU= ',MU,'MU2= ',MU2
        WRITE (44,420) 'UPPER_PERCENTILE= ',UPPER_PERCENTILE,
1      'LOWER_PERCENTILE= ',LOWER_PERCENTILE
        WRITE (44,410) 'MU1= ',MU1,'NU= ',NU,'MUPRIM= ',MUPRIM

        DO 110 J=NU1,N
          WRITE (44,440) 'R(',J,')'
          WRITE (44,490) (R(I,J),I=1,M)
110      CONTINUE
        WRITE (44,400) 'GOING INTO DO LOOP 130'
      END IF

      DO 130 I=1,M
        RMEAN(I)=0.
        DO 120 J=NU1,NU2
          RMEAN(I)=RMEAN(I)+R(I,J)
120      CONTINUE
        RMEAN(I)=RMEAN(I)/NU
130      CONTINUE

      IF (IDG.EQ.1.OR.IDG.EQ.2) THEN
        WRITE (44,400) 'FORMAT OF WRITE IS I,RMEAN'
        WRITE (44,460) (I,RMEAN(I),I=1,M)
        WRITE (44,400) 'GOING INTO DO LOOP 150'
      END IF

```



```

DO 150 I=1,M
  DO 140 J=NU1,N
    DR(I,J)=R(I,J)-RMEAN(I)
140    CONTINUE
150    CONTINUE

IF (IDG.EQ.1.OR.IDG.EQ.2) THEN
  DO 160 J=NU1,N
    WRITE (44,440) 'DR(',J,')'
    WRITE (44,490) (DR(I,J),I=1,M)
160    CONTINUE
    WRITE (44,400) 'GOING INTO DO LOOP 190'
  END IF

DO 190 K=1,MU
  DO 180 Q=1,MU
    A(K,Q)=0.
    DO 170 J=NU1,NU2
      A(K,Q)=A(K,Q)+DR(MU1+K-1,J)*DR(MU1+Q-1,J)
170    CONTINUE
180    CONTINUE
190    CONTINUE

IF (IDG.EQ.1.OR.IDG.EQ.2) THEN
  DO 200 Q=1,MU
    WRITE (44,440) 'A(',Q,')'
    WRITE (44,510) (A(K,Q),K=1,MU)
200    CONTINUE
    WRITE (44,400) 'GOING INTO DO LOOP 230'
  END IF

DO 230 P=1,MUPRIM
  DO 220 Q=1,MU
    B(P,Q)=0.
    DO 210 J=NU1,NU2
      B(P,Q)=B(P,Q)+DR(MU2+P,J)*DR(MU1+Q-1,J)
210    CONTINUE
220    CONTINUE
230    CONTINUE

IF (IDG.EQ.1.OR.IDG.EQ.2) THEN
  DO 240 Q=1,MU
    WRITE (44,440) 'B(',Q,')'
    WRITE (44,510) (B(P,Q),P=1,MUPRIM)
240    CONTINUE
    WRITE (44,400) 'GOING INTO DO LOOP 260'
  END IF

K=0

DO 260 J=1,MU
  DO 250 I=1,MU
    K=K+1
    APRIME(K)=A(I,J)

```

```

250     CONTINUE
260     CONTINUE

IF (IDG.EQ.1.OR.IDG.EQ.2) THEN
    WRITE (44,400) 'FORMAT OF WRITE IS K,APRIME'
    WRITE (44,460) (I,APRIME(I),I=1,K)
END IF

DETA(1)=3.
DETA(2)=0.

CALL GJR (A,MU,MU,MU,MU,*390,JDUM,DETA)

IF (DETA(1).EQ.0.) GO TO 390

K=MU*MU+1

DO 280 J=MU,1,-1
    DO 270 I=MU,1,-1
        K=K-1
        AINV(I,J)=APRIME(K)
270     CONTINUE
280     CONTINUE

IF (IDG.EQ.1.OR.IDG.EQ.2) THEN
    DO 290 J=1,MU
        WRITE (44,450) 'AINV(',J,')'
        WRITE (44,500) (AINV(P,J),P=1,MU)
290     CONTINUE
    WRITE (44,400) 'GOING INTO DO LOOP 320'
END IF

DO 320 P=1,MUPRIM
    DO 310 K=1,MU
        C(P,K)=0.
        DO 300 Q=1,MU
            C(P,K)=C(P,K)+B(P,Q)*AINV(Q,K)
300     CONTINUE
310     CONTINUE
320     CONTINUE

IF (IDG.EQ.1.OR.IDG.EQ.2) THEN
    DO 330 Q=1,MU
        WRITE (44,440) 'C(',Q,')'
        WRITE (44,500) (C(P,Q),P=1,MUPRIM)
330     CONTINUE
    WRITE (44,400) 'GOING INTO DO LOOP 360'
END IF

DO 360 P=1,M
    IF (P.GT.MU2) GO TO 340
    RPRED(P)=R(P,NU2+1)
    GO TO 360
340     RPRED(P)=0.

```

```

DO 350 K=1,MU
    RPRED(P)=RPRED(P)+C(P-MU2,K)*DR(MU1+K-1,NU2+1)
350    CONTINUE
    RPRED(P)=RPRED(P)+RMEAN(P)
360    CONTINUE

IF (IDG.EQ.1.OR.IDG.EQ.2) THEN
    WRITE (44,400) 'FORMAT OF WRITE IS I,RPRED'
    WRITE (44,460) (I,RPRED(I),I=1,M)
END IF

MU2P1=MU2+1

IF (IDG.EQ.1.OR.IDG.EQ.2) THEN
    WRITE (44,430) 'MU2P1= ',MU2P1
END IF

C*****
C WHEN I2=3 THE SUBROUTINE BYPASSES THE SIGMA TO CALCULATE
C 95% AND 5% USING EMPIRICAL STATISTICS WITH QUANTILE FUNCTION.
C*****
C

IF (I2.EQ.3) THEN

    CALL PRCNTILE (DR,C,NU2)

    IF (IDG.EQ.1.OR.IDG.EQ.2) THEN
        WRITE (44,410) 'M= ',M,'N= ',N,'NU1= ',NU1
        WRITE (44,410) 'NU2= ',NU2,'MU= ',MU,'MU2= ',MU2
        WRITE (44,420) 'UPPER_PERCENTILE= ',UPPER_PERCENTILE,
1          'LOWER_PERCENTILE= ',LOWER_PERCENTILE
        WRITE (44,410) 'MU1= ',MU1,'NU= ',NU,'MUPRIM= ',MUPRIM
        WRITE (44,400) 'FORMAT OF WRITE IS I,RMEAN'
        WRITE (44,460) (I,RMEAN(I),I=1,2*M)
        WRITE (44,400) 'FORMAT OF WRITE IS I,RPRED'
        WRITE (44,460) (I,RPRED(I),I=1,2*M)
        WRITE (44,400) 'FORMAT OF WRITE IS I,ERRUP'
        WRITE (44,460) (I,ERRUP(I),I=1,2*M)
        WRITE (44,400) 'FORMAT OF WRITE IS I,ERRDN'
        WRITE (44,460) (I,ERRDN(I),I=1,2*M)
        WRITE (44,400) 'YOU ARE EXITING LINCMC SUBROUTINE'
    END IF

C *****
C *      RETURN STATEMENT TO SEND BACK TO MAIN      *
C *****

RETURN

ELSE

    IF (IDG.EQ.1.OR.IDG.EQ.2) THEN
        WRITE (44,400) 'GOING INTO DO LOOP 380'
    
```

```

        WRITE (44,400) 'FORMAT OF WRITE IS P,SIGMA'
END IF

DO 380 P=MU2P1,M
    SIGMA=0.
    DO 370 J=NU1,NU2
        SIGMA=SIGMA+(R(P,J)-RPRED(P))**2
370    CONTINUE
        SIGMA=SQRT(SIGMA/(NU-MU+1))
        ERRUP(P)=RPRED(P)+2.*SIGMA
        ERRDN(P)=RPRED(P)-2.*SIGMA
        IF (IDG.EQ.1.OR.IDG.EQ.2) THEN
            WRITE (44,480) P,SIGMA
        END IF
        IF (ERRDN(P).LT.0.) ERRDN(P)=0.
380    CONTINUE

    IF (IDG.EQ.1.OR.IDG.EQ.2) THEN
        WRITE (44,470) (P,ERRUP(P),ERRDN(P),P=MU2P1,M)
        WRITE (44,400) 'IF I2 NOT EQUAL TO 3 RETURN TO MAIN'
    END IF

END IF

RETURN

390  WRITE (6,*) 'MATRIX WAS SINGLR'
    STOP

400  FORMAT (A50)
410  FORMAT (2X,A7,I4,1X,A7,I4,1X,A7,I4)
420  FORMAT (2X,A18,F5.2,1X,A18,F5.2)
430  FORMAT (2X,A7,I5)
440  FORMAT (2X,A3,I3,A2)
450  FORMAT (2X,A6,I3,A2)
460  FORMAT (6(1X,I3,1X,F8.2))
470  FORMAT (3(1X,I3,1X,F8.2,1X,F8.2))
480  FORMAT (2X,I4,1X,F8.2)
490  FORMAT (12F6.1)
500  FORMAT (5(1X,F9.6))
510  FORMAT (8(1X,F8.2))

```

END

SUBROUTINE QUANTILE (DDR,NU2,P)

```

C-----
C  THIS SUBROUTINE RANKS THE DATA IN ASCENDING ORDER THEN ALLOCATES
C  THE QUANTILE DISTRIBUTION FOR THE DATA GIVEN.
C-----
C
C-----
C  THE DATA IS DIMENSIONED STARTING AT ZERO WHEN THE DIMENSION

```

```

C      STATEMENT HAS 0: IN IT.
C-----

      INTEGER*2 NU2,P,K,IDG,I,M,MU2P1,MU2,MU1,MU,NU1,NU,I2
      REAL*8 DDR(50),RNKANS(50),C,QUANT(50)

      COMMON /BLK1/ I2
      COMMON /IDG/ IDG
      COMMON /ANS2/ RNKANS,QUANT
      COMMON /COUNT/ NU,NU1,MU,MU1,MU2,MU2P1,M

      IF(IDG.EQ.1.OR.IDG.EQ.2) THEN
      WRITE (44,120) 'NU2= ',NU2
      WRITE (44,110) 'WRITE FORMAT IS I,DDR'
      WRITE (44,140) (I,DDR(I),I=1,NU2)
      WRITE (44,110) 'CALL RANKIT'
      END IF

C-----
C  CALLS RANKIT SUBROUTINE TO SET THE DATA IN ASCENDING ORDER
C-----

      CALL RANKIT (DDR,NU2,RNKANS)

      IF(IDG.EQ.1.OR.IDG.EQ.2) THEN
      WRITE (44,110) 'WRITE FORMAT IS I,RNKANS'
      WRITE (44,140) (I,RNKANS(I),I=1,NU2)
      END IF

C-----
C  CALCULATE THE QUANTILE FUNCTION SEE APPENDIC C OF TM TBD FOR
C  EQUATION.
C-----

      DO I=1,NU2
C      QUANT(I)=(FLOAT(I)-.5)/(FLOAT(NU2))
      QUANT(I)=FLOAT(I)/(FLOAT(NU2)+1.0)
      END DO

C-----WRITE STAEMENTS FOR TROUBLE SHOOTING PROGRAM-----

      IF (IDG.EQ.1.OR.IDG.EQ.2) THEN
      WRITE (44,120) 'P= ',P
      WRITE (44,110) 'FORMAT OF WRITE IS K,RNKANS,QUANT'
      WRITE (44,130) (K,RNKANS(K),QUANT(K),K=1,NU2)
C      WRITE(44,6000) 'CALL FNDLAMBDA'
      END IF

      RETURN

C-----FORMAT STATEMENTS-----
110  FORMAT (A50)
120  FORMAT (2X,A7,I5)
130  FORMAT (1X,I3,1X,F8.2,F8.2)

```

```
140  FORMAT (6(1X,I3,1X,F8.2))  
      END
```

APPENDIX E

MSFC Monthly Solar Activity Memorandum Example

EL23 (Example)

TO: Distribution

FROM: EL23/Chief, Electromagnetics and Aerospace Environments Branch

SUBJECT: Solar Activity Inputs for Upper Atmospheric Models Used in Programs
to Estimate Spacecraft Orbital Lifetime

The solar activity information in this memorandum is provided as input data for upper atmospheric models to insure compatibility between calculations made for spacecraft orbital lifetime predictions. The Marshall Engineering Thermosphere Model (NASA CR-179359 and CR-179389), as well as the NASA/MSFC Global Reference Atmospheric Model-1995 Version (NASA TM-4715), require inputs of the 10.7-cm solar flux ($F_{10.7}$) and the geomagnetic index (A_p) to compute atmospheric density. Statistical estimates are provided for the future 13-month smoothed values of these parameters.

To provide a better statistical estimate of the forthcoming solar minimum and the temporal profile for cycle 23, the MSFC linear regression program has been revised. The MSFC linear regression program is based on the Lagrangian least-squares statistical technique. This technique is described in the papers "Lagrangian Least-Squares Prediction of Solar Flux ($\bar{F}_{10.7}$)", Journal of Geophysical Research, Vol. 89, Number A1, Pages 11 through 16, January 1, 1984. A more detailed explanation of the MSFC linear regression computer program and modifications that took place in 1995 and 1996 are in the paper "Statistical Technique for Intermediate and Long-Range Estimation of 13-Month Smoothed Solar Flux and Geomagnetic Index" (NASA TM-TBD) which is being processed for publication.

MSFC's linear regression program uses the 13-month smoothed 10.7-cm solar flux ($\bar{F}_{10.7}$) cycles 1 through 21 converted and observed data base and the 13-month smoothed geomagnetic index (\bar{A}_p) cycles 13 through 21 converted and observed data base. The program estimates the intermediate-term (months) and long-term (years) behavior of the 13-month smoothed solar flux ($\bar{F}_{10.7}$) and 13-month smoothed geomagnetic index (\bar{A}_p), up to 132 months into the future, initialized from the cycle 22 June 1989 maximum. Once cycle 23 has been established from observed solar activity data, the solar flux program and geomagnetic index program will be re-initialized at the minimum for cycle 23. The mean cycle period of 11 years (132 months) will be assumed for cycle 23. Cycle 24 will then be estimated using the solar flux statistics for cycles 1 through 22 and geomagnetic index statistics for cycles 13 through 22.

The changes of orbital density associated with short-term (days) variations in the daily 10.7-cm solar flux and the 3-hourly geomagnetic index, required as inputs to the upper atmospheric models, are not represented by the 13-month smoothed statistical estimates

given in these tables. This dynamic component of the total atmospheric density cannot be estimated into the future with any acceptable degree of statistical confidence using existing techniques. Representative data sets, based on past daily 10.7-cm solar flux and 3-hourly geomagnetic index values, may be utilized to compute this dynamic component of the orbital altitude density.

The design requirements for solar activity and associated on-orbit density values are specified in the appropriate spacecraft and space vehicle project design requirements documentation. These documents should be consulted for this information.

Enclosed are the following tables and figures:

Table E-1 (Enclosure 1) contains values for the recent monthly 10.7-cm solar flux and geomagnetic index.

Table E-2 (Enclosure 2) contains the mid-point calculated values of the 13-month smoothed observed 10.7-cm solar flux.

Table E-3 (Enclosure 3) contains the statistical estimate of 13-month smoothed 10.7-cm solar flux and geomagnetic index for balance of cycle 22, cycle 23, and the beginning of cycle 24.

Figure E-1 (Enclosure 4) is a plot of solar cycle 22 monthly mean and 13-month smoothed observed 10.7-cm solar flux.

Figure E-2 (Enclosure 4) is a plot of the 13-month smoothed 10.7-cm solar flux statistical estimates for solar cycles 22, 23, and beginning of cycle 24.

The 50 percentile values in table E-3 and figure E-2, at and beyond maximum of cycle 23, are approximated by the arithmetic mean and given with 95 percentile and 5 percentile values.

The information in this memorandum is based upon data received from the National Research Council of Canada for the Series C 10.7-cm solar flux data and the Institute for Geophysics in Gottingen, Germany for the geomagnetic index data.

Questions on the contents of this memorandum may be addressed to Harold Euler at (205) 544-2282.

APPROVAL:

4 Enclosures

TABLE E-1
RECENT MONTHLY MEAN SOLAR ACTIVITY VALUES

		Solar Flux $F_{10.7}$ (Series C)	Geomagnetic Index A_p
1994	January	115.0	15
	February	99.6	30
	March	90.4	24
	April	79.1	29
	May	79.9	26
	June	77.3	14
	July	80.5	11
	August	76.1	8
	September	79.1	8
	October	87.7	22
	November	80.9	14
	December	84.1	13
1995	January	82.7	11
	February	85.6	15
	March	85.1	15
	April	77.7	16
	May	75.6	18
	June	75.7	11
	July	73.8	8
	August	73.8	9
	September	72.0	13
	October	77.9	16
	November	74.2	9
	December	72.6	9
1996	January	74.5	9
	February	71.5	10
	March	70.7	11
	April	69.1*	12*
	May	69.7*	7*
	June	69.5*	6*

Solar flux in units of 10^4 JANSKY (where one JANSKY equals 10^{-26} W m⁻² Hz⁻¹ Bandwidth)

* Provisional

TABLE E-2
13-MONTH SMOOTHED 10.7-CM SOLAR FLUX

MONTH	1989	1990	1991	1992	1993	1994	1995
JANUARY	190.2	200.4	205.5	181.8	125.6	92.7	80.6
FEBRUARY	194.1	200.5	206.3	174.8	123.0	91.2	80.2
MARCH	199.7	198.7	205.9	168.5	120.6	90.2	79.9
APRIL	204.4	195.6	206.8	162.9	118.1	89.3	79.2
MAY	209.3	192.4	207.1	158.8	114.8	88.2	78.5
JUNE	213.1	189.9	207.4	154.2	111.3	86.7	77.7
JULY	212.6	190.4	207.7	146.6	109.6	84.5	76.9
AUGUST	209.7	193.9	206.8	138.9	107.6	82.5	76.0
SEPTEMBER	207.2	198.3	203.9	133.7	103.9	81.7	74.8
OCTOBER	206.3	200.6	199.7	130.5	100.4	81.4	73.8
NOVEMBER	206.1	201.2	195.4	128.2	97.5	81.2	73.2*
DECEMBER	203.3	202.7	188.9	127.3	94.8	81.0	72.7*

*Provisional

NOTE: TABLE E-2 contains the 13-month smoothed 10.7-cm solar flux at the mid-point computed from the National Research Council of Canada, Ottawa and Penticton Series C observed monthly values.

TABLE E-3 ESTIMATES OF 13-MONTH SMOOTHED SOLAR ACTIVITY FOR
BALANCE OF CYCLE 22, CYCLE 23, AND BEGINNING OF CYCLE 24

TIME		10.7-CM SOLAR FLUX PERCENTILE			GEOMAGNETIC INDEX PERCENTILE		
		95.0%	50%	($\bar{F}_{10.7}$) 5.0%	95.0%	50%	(\bar{A}_p) 5.0%
1996.0003	JAN	74.1	72.6	71.3	10.3	10.1	9.7
1996.0837	FEB	75.1	72.5	70.2	10.8	10.1	9.3
1996.1670	MAR	76.5	72.6	69.5	11.3	10.2	8.8
1996.2503	APR	78.3	72.8	68.5	11.8	10.2	8.4
1996.3337	MAY	80.1	73.0	67.6	12.1	10.2	8.0
1996.4170	JUN	82.4	73.3	66.8	12.5	10.2	7.8
1996.5003	JUL	85.1	73.7	66.2	12.8	10.1	7.7
1996.5837	AUG	87.8	74.2	65.2	13.0	10.2	7.7
1996.6670	SEP	91.2	74.8	64.3	13.1	10.2	7.9
1996.7503	OCT	96.0	75.7	63.4	13.0	10.2	8.1
1996.8337	NOV	100.9	76.9	62.5	13.5	10.3	8.4
1996.9170	DEC	105.4	78.2	61.7	14.5	10.4	8.8
1997.0003	JAN	109.3	79.5	60.9	15.0	10.5	8.8
1997.0837	FEB	112.3	80.7	60.5	15.3	10.6	8.7
1997.1670	MAR	115.4	82.1	60.4	15.7	10.8	8.7
1997.2503	APR	119.4	83.5	60.2	15.9	11.0	8.5
1997.3337	MAY	123.7	85.1	60.3	15.9	11.1	8.5
1997.4170	JUN	128.4	86.7	60.6	16.2	11.1	8.5
1997.5003	JUL	134.1	88.6	60.8	16.4	11.2	8.7
1997.5837	AUG	140.1	90.8	61.3	16.4	11.4	8.9
1997.6670	SEP	144.8	93.0	62.3	16.5	11.7	9.1
1997.7503	OCT	148.1	95.1	63.3	16.8	12.0	9.5
1997.8337	NOV	150.1	97.2	64.9	16.9	12.4	10.0
1997.9170	DEC	152.1	99.2	66.8	16.4	12.7	10.6
1998.0003	JAN	154.9	101.4	68.3	15.8	12.8	10.9
1998.0837	FEB	158.2	103.7	69.6	15.6	12.9	10.4
1998.1670	MAR	161.5	105.9	71.4	15.7	13.1	10.4
1998.2503	APR	165.1	108.2	73.1	15.7	13.3	10.4
1998.3337	MAY	170.3	110.6	73.7	16.0	13.4	10.4
1998.4170	JUN	175.6	113.0	74.3	16.3	13.5	10.4
1998.5003	JUL	178.8	115.5	76.9	16.5	13.5	10.7
1998.5837	AUG	180.4	118.0	78.5	16.4	13.5	11.0
1998.6670	SEP	182.0	120.4	80.0	16.2	13.3	11.2
1998.7503	OCT	183.6	122.9	81.6	16.3	13.3	11.2
1998.8337	NOV	185.1	125.2	83.1	16.4	13.3	11.3
1998.9170	DEC	186.6	127.6	84.6	16.5	13.2	11.4
1999.0003	JAN	188.1	129.9	86.0	16.6	13.2	11.5
1999.0837	FEB	189.6	132.2	87.5	16.7	13.2	11.5
1999.1670	MAR	191.0	134.3	88.9	16.8	13.2	11.6
1999.2503	APR	192.3	136.4	90.2	16.9	13.2	11.7
1999.3337	MAY	193.6	138.4	91.5	17.0	13.1	11.8
1999.4170	JUN	194.9	140.3	92.7	17.1	13.1	11.8
1999.5003	JUL	196.0	142.1	93.8	17.2	13.1	11.9
1999.5837	AUG	197.1	143.8	94.9	17.3	13.1	11.9
1999.6670	SEP	198.1	145.3	95.8	17.3	13.1	12.0
1999.7503	OCT	199.0	146.7	96.7	17.4	13.1	12.0
1999.8337	NOV	199.8	147.9	97.5	17.5	13.0	12.1
1999.9170	DEC	200.5	149.0	98.2	17.5	13.0	12.1
2000.0003	JAN	201.1	149.9	98.8	17.5	13.0	12.1
2000.0837	FEB	201.5	150.6	99.2	17.6	13.0	12.2

TABLE E-3 ESTIMATES OF 13-MONTH SMOOTHED SOLAR ACTIVITY FOR
BALANCE OF CYCLE 22, CYCLE 23, AND BEGINNING OF CYCLE 24

TIME		10.7-CM SOLAR FLUX PERCENTILE			GEOMAGNETIC INDEX PERCENTILE		
		95.0%	50%	($\bar{F}_{10.7}$) 5.0%	95.0%	50%	(\bar{A}_p) 5.0%
2000.1670	MAR	201.9	151.2	99.6	17.6	13.0	12.2
2000.2503	APR	202.1	151.5	99.8	17.6	13.0	12.2
2000.3337	MAY	202.2	151.6	99.8	17.6	13.0	12.2
2000.4170	JUN	239.0	158.1	98.5	20.0	15.2	11.9
2000.5003	JUL	234.3	156.1	97.8	19.7	15.2	12.2
2000.5837	AUG	230.8	154.2	96.7	19.3	15.3	12.2
2000.6670	SEP	230.3	152.7	95.3	19.0	15.5	12.3
2000.7503	OCT	230.0	151.1	93.7	18.8	15.6	12.5
2000.8337	NOV	227.8	149.7	92.5	18.6	15.6	12.1
2000.9170	DEC	225.6	148.1	91.6	18.6	15.6	11.5
2001.0003	JAN	224.8	146.6	90.8	18.5	15.6	11.4
2001.0837	FEB	223.7	144.9	90.0	18.6	15.6	11.4
2001.1670	MAR	222.6	143.0	89.0	18.8	15.6	11.4
2001.2503	APR	219.3	140.8	88.3	19.3	15.7	11.4
2001.3337	MAY	214.8	138.7	87.4	19.9	15.8	11.5
2001.4170	JUN	211.2	137.0	86.0	20.8	16.0	11.5
2001.5003	JUL	205.9	135.8	85.3	21.3	16.2	11.4
2001.5837	AUG	200.9	134.7	84.7	21.1	16.3	11.3
2001.6670	SEP	196.3	133.7	83.7	20.8	16.2	11.1
2001.7503	OCT	191.4	132.8	82.4	21.3	16.3	11.1
2001.8337	NOV	187.0	131.4	81.0	22.4	16.5	11.2
2001.9170	DEC	182.3	129.7	80.1	22.6	16.7	11.7
2002.0003	JAN	177.9	127.8	79.5	22.2	16.9	12.4
2002.0837	FEB	176.3	126.0	78.8	22.0	16.8	12.5
2002.1670	MAR	176.1	124.4	78.0	22.1	16.8	12.6
2002.2503	APR	176.7	123.0	77.1	22.7	17.0	12.7
2002.3337	MAY	177.8	121.9	76.1	23.3	17.3	12.8
2002.4170	JUN	177.3	120.5	75.2	23.3	17.3	13.2
2002.5003	JUL	175.1	118.6	74.3	23.3	17.2	13.9
2002.5837	AUG	171.4	116.7	73.3	22.8	17.2	14.1
2002.6670	SEP	166.1	114.8	72.3	21.2	17.0	13.7
2002.7503	OCT	161.6	113.2	71.8	21.3	16.7	13.2
2002.8337	NOV	160.0	111.8	71.5	21.2	16.7	13.1
2002.9170	DEC	159.0	110.3	71.1	21.0	16.6	13.0
2003.0003	JAN	157.4	108.9	70.9	20.5	16.4	12.8
2003.0837	FEB	154.7	107.4	71.0	20.0	16.2	12.6
2003.1670	MAR	150.4	105.6	70.9	19.4	15.8	12.3
2003.2503	APR	144.8	103.9	70.7	18.7	15.6	12.3
2003.3337	MAY	139.0	102.5	70.5	17.9	15.3	12.2
2003.4170	JUN	133.3	101.3	70.4	17.6	15.1	12.2
2003.5003	JUL	128.5	100.1	70.3	17.9	14.9	12.1
2003.5837	AUG	124.6	98.8	70.6	18.3	14.7	11.9
2003.6670	SEP	121.5	97.6	71.0	18.5	14.6	12.0
2003.7503	OCT	119.3	96.6	71.1	19.1	14.6	11.5
2003.8337	NOV	117.9	95.5	71.3	19.6	14.6	11.1
2003.9170	DEC	116.4	94.6	71.1	20.0	14.8	11.4
2004.0003	JAN	117.6	93.5	70.6	20.4	15.0	11.4
2004.0837	FEB	117.2	92.4	70.3	20.6	15.1	11.4
2004.1670	MAR	116.3	91.6	69.9	20.9	15.4	11.6
2004.2503	APR	116.8	90.9	69.5	21.6	15.6	11.7

TABLE E-3 ESTIMATES OF 13-MONTH SMOOTHED SOLAR ACTIVITY FOR
BALANCE OF CYCLE 22, CYCLE 23, AND BEGINNING OF CYCLE 24

TIME		10.7-CM SOLAR FLUX PERCENTILE			GEOMAGNETIC INDEX PERCENTILE		
		95.0%	50%	($\bar{F}_{10.7}$) 5.0%	95.0%	50%	(\bar{A}_p) 5.0%
2004.3337	MAY	117.3	89.9	69.2	21.8	15.8	11.7
2004.4170	JUN	117.6	89.0	68.9	21.8	16.0	11.6
2004.5003	JUL	117.4	88.0	68.6	22.0	16.3	11.6
2004.5837	AUG	116.7	86.9	68.2	22.1	16.6	11.9
2004.6670	SEP	115.5	85.9	67.9	22.4	16.9	11.9
2004.7503	OCT	113.9	85.1	67.7	23.0	17.0	11.9
2004.8337	NOV	110.7	84.1	67.6	23.6	17.0	12.4
2004.9170	DEC	105.3	83.0	67.5	24.1	16.9	12.8
2005.0003	JAN	99.5	82.0	67.5	24.3	16.8	12.8
2005.0837	FEB	98.3	81.0	67.3	24.1	16.5	12.7
2005.1670	MAR	98.8	80.2	67.2	23.4	16.0	12.8
2005.2503	APR	99.9	79.7	67.2	22.5	15.6	12.8
2005.3337	MAY	100.4	79.2	67.2	21.7	15.3	12.8
2005.4170	JUN	98.8	78.6	67.2	21.4	14.9	12.4
2005.5003	JUL	95.8	78.1	67.1	20.9	14.5	11.8
2005.5837	AUG	93.2	77.6	67.1	20.5	14.1	11.2
2005.6670	SEP	91.7	77.1	67.0	19.9	13.6	10.5
2005.7503	OCT	91.6	76.5	67.0	19.0	13.1	9.9
2005.8337	NOV	91.1	76.0	67.0	17.9	12.7	9.4
2005.9170	DEC	90.7	75.6	67.0	16.8	12.3	8.9
2006.0003	JAN	90.0	75.1	67.0	16.2	12.0	8.5
2006.0837	FEB	89.1	74.7	67.1	16.1	11.8	8.3
2006.1670	MAR	88.4	74.4	67.3	16.4	11.8	8.0
2006.2503	APR	87.5	74.1	67.3	16.6	11.6	7.6
2006.3337	MAY	86.1	73.7	67.4	16.3	11.4	7.3
2006.4170	JUN	84.6	73.3	67.5	16.1	11.3	7.2
2006.5003	JUL	82.5	72.9	67.6	15.9	11.3	7.1
2006.5837	AUG	80.0	72.5	67.7	15.4	11.2	7.2
2006.6670	SEP	77.7	72.1	68.0	15.2	11.2	7.4
2006.7503	OCT	77.4	71.9	68.0	15.3	11.2	7.6
2006.8337	NOV	77.0	71.7	68.0	15.8	11.2	7.8
2006.9170	DEC	77.5	71.6	68.0	16.0	11.3	7.9
2007.0003	JAN	78.4	71.5	67.9	16.2	11.3	7.9
2007.0837	FEB	80.0	71.6	67.9	16.4	11.3	8.1
2007.1670	MAR	82.1	71.8	67.7	16.5	11.3	8.1
2007.2503	APR	84.0	72.0	67.6	16.8	11.2	8.1
2007.3337	MAY	86.4	72.3	67.7	17.0	11.2	8.1
2007.4170	JUN	89.3	72.7	67.6	17.1	11.1	8.2
2007.5003	JUL	92.3	73.0	67.6	17.1	11.1	8.2
2007.5837	AUG	95.8	73.5	67.6	16.8	11.0	8.1
2007.6670	SEP	101.2	74.2	67.9	16.5	11.0	8.0
2007.7503	OCT	106.7	75.2	68.1	15.7	10.9	8.0
2007.8337	NOV	112.0	76.3	68.5	15.0	10.9	8.0
2007.9170	DEC	116.6	77.3	68.5	15.5	10.9	8.2
2008.0003	JAN	120.1	78.5	68.8	15.5	10.9	8.2
2008.0837	FEB	123.6	79.7	69.1	15.7	11.0	8.3
2008.1670	MAR	127.8	81.0	69.3	15.9	11.1	8.2
2008.2503	APR	132.5	82.5	69.5	16.0	11.2	8.3
2008.3337	MAY	137.4	84.1	70.2	16.2	11.2	8.4
2008.4170	JUN	143.4	85.9	71.0	16.4	11.3	8.6

TABLE E-3 ESTIMATES OF 13-MONTH SMOOTHED SOLAR ACTIVITY FOR
BALANCE OF CYCLE 22, CYCLE 23, AND BEGINNING OF CYCLE 24

TIME		10.7-CM SOLAR FLUX PERCENTILE			GEOMAGNETIC INDEX PERCENTILE		
		95.0%	50%	($\bar{F}_{10.7}$) 5.0%	95.0%	50%	(\bar{A}_p) 5.0%
2008.5003	JUL	149.7	87.9	71.8	16.4	11.5	8.9
2008.5837	AUG	154.6	90.1	72.5	16.5	11.7	9.1
2008.6670	SEP	158.2	92.2	73.7	16.8	12.0	9.6
2008.7503	OCT	160.3	94.2	74.2	16.8	12.3	10.1
2008.8337	NOV	162.1	96.2	74.4	16.3	12.5	10.4
2008.9170	DEC	164.4	98.6	74.7	15.8	12.7	10.4
2009.0003	JAN	167.7	100.9	74.9	15.6	12.9	10.3
2009.0837	FEB	171.4	103.0	75.0	15.7	13.1	10.4
2009.1670	MAR	174.9	105.3	74.9	15.8	13.3	10.5
2009.2503	APR	179.5	107.5	74.8	16.2	13.4	10.6
2009.3337	MAY	185.0	109.9	74.7	16.5	13.5	10.7
2009.4170	JUN	188.6	112.2	75.0	16.9	13.6	11.2
2009.5003	JUL	190.3	114.7	75.3	17.4	13.7	12.1
2009.5837	AUG	190.7	117.3	75.6	17.7	13.7	12.0
2009.6670	SEP	191.0	119.6	76.6	17.7	13.6	11.4
2009.7503	OCT	193.5	121.6	77.4	17.9	13.6	10.9
2009.8337	NOV	196.6	123.2	77.7	18.2	13.7	10.6
2009.9170	DEC	199.1	124.9	78.7	17.9	13.7	10.5
2010.0003	JAN	202.7	126.8	80.0	17.3	13.7	10.9
2010.0837	FEB	208.4	128.8	81.5	17.6	13.8	11.0
2010.1670	MAR	212.9	131.0	83.5	18.4	13.8	10.5
2010.2503	APR	215.1	133.1	84.8	18.6	13.9	10.5
2010.3337	MAY	218.6	135.5	85.8	19.1	14.2	10.7
2010.4170	JUN	223.7	138.3	87.7	19.9	14.4	10.9
2010.5003	JUL	226.5	141.0	89.5	19.6	14.3	11.0
2010.5837	AUG	228.2	143.2	90.8	19.7	14.2	10.7
2010.6670	SEP	230.2	145.6	93.2	19.9	14.3	10.8
2010.7503	OCT	232.2	148.2	95.7	20.3	14.3	10.5
2010.8337	NOV	235.2	150.7	97.5	20.5	14.4	10.4
2010.9170	DEC	238.5	152.9	98.7	20.5	14.6	11.0
2011.0003	JAN	240.7	154.8	98.7	20.8	14.8	11.4
2011.0837	FEB	240.1	156.6	98.1	21.1	15.0	11.4
2011.1670	MAR	239.8	158.8	97.7	21.6	15.2	11.5
2011.2503	APR	241.6	161.0	99.1	22.1	15.5	11.7
2011.3337	MAY	241.6	161.0	99.1	22.1	15.5	11.7

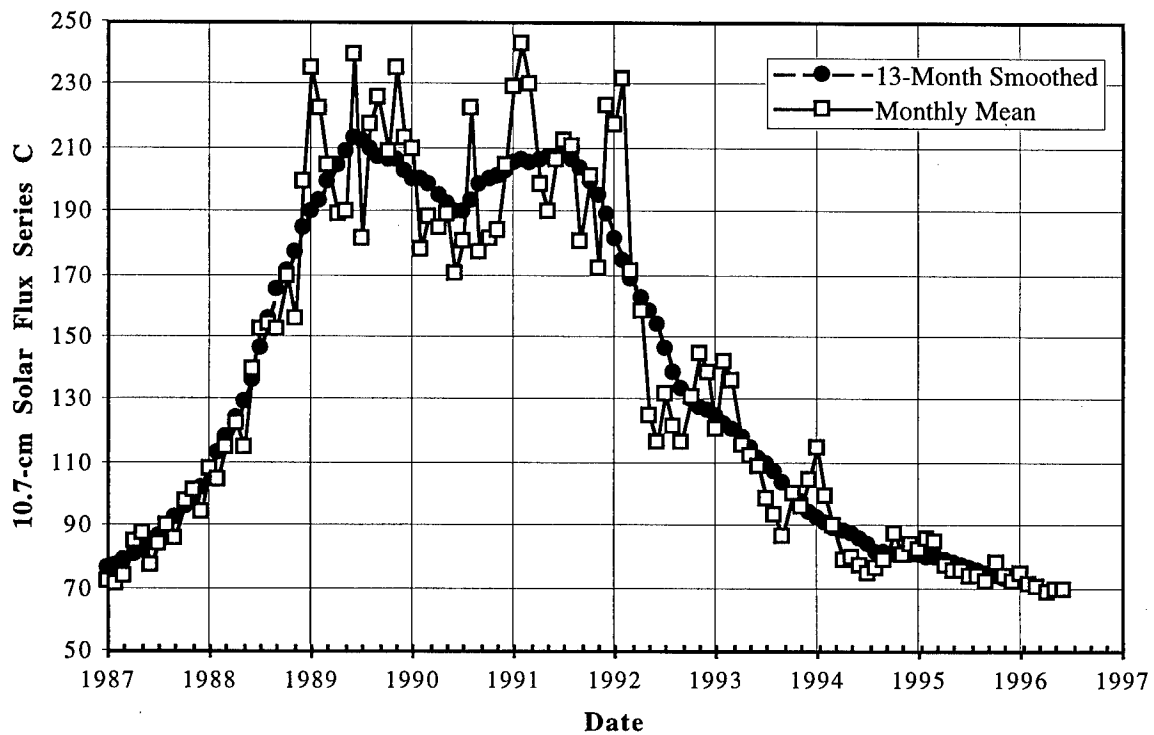


Figure E-1. Plot of Recent Monthly Mean and 13-Month Smoothed Solar Flux

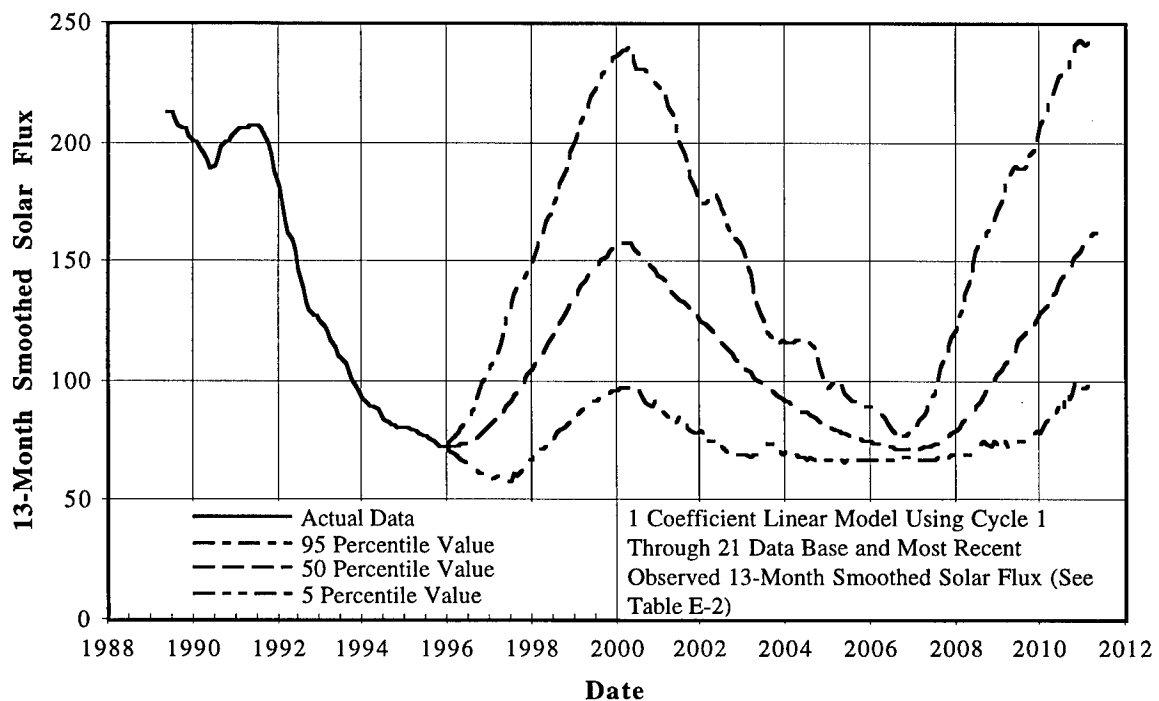


Figure E-2. Estimate of 13-Month Smoothed Solar Flux for Balance of Cycle 22, 23, and Beginning of Cycle 24

APPENDIX F

MSFC Lagrangian Linear Regression Technique Historical Plots

MSFC Lagrangian Linear Regression Technique (MLLRT) Historical Plots were developed to provide an assessment of the MLLRT over the last three cycles. The plots are made starting on the month and year specified on each plot. These plots are made for one month for each year during cycles 20, 21, and 22. The plots show the estimates as in the MSFC Monthly Solar Activity memorandum with the observed 13-Month Smoothed $F_{10.7}$ added. Each estimate plot is for five years. This appendix contains also the completed cycle plots for the minimum to minimum cycles and the maximum to maximum cycles. Concerning the maximum to maximum cycle plots, the cycles are identified by cycle number at maximum initialization date. For example, maximum to maximum cycle 20 starts on July 1970 and finishes at the maximum estimate of cycle 21. Concerning the minimum to minimum cycle plots, the cycles are identified by cycle number at minimum initialization date, e. g., minimum to minimum cycle 20 starts on October 1964 and finishes at the minimum estimate of cycle 21.

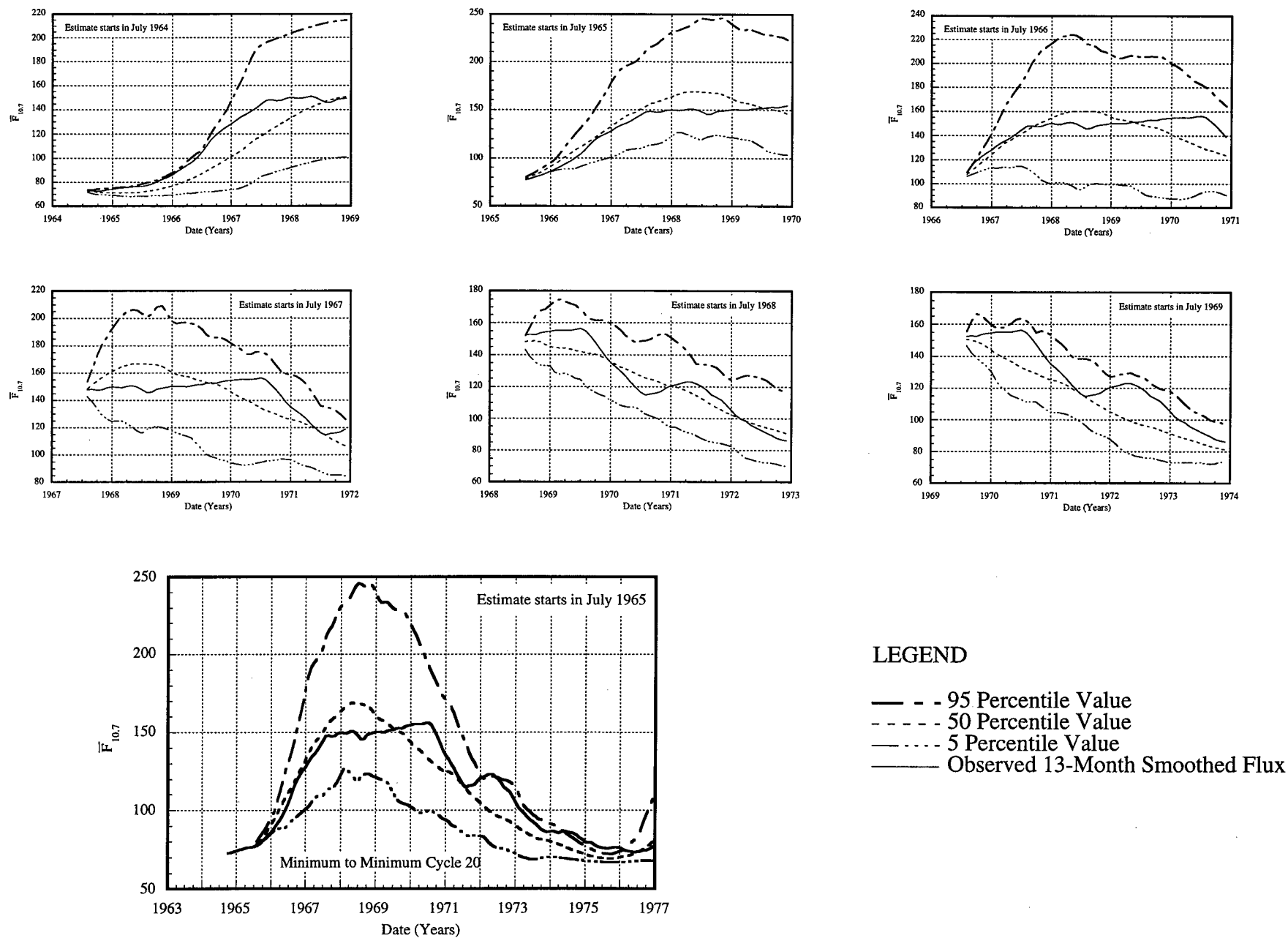


Figure F-1. Minimum to Minimum Cycle 20 MLLRT Model Results Evaluation.

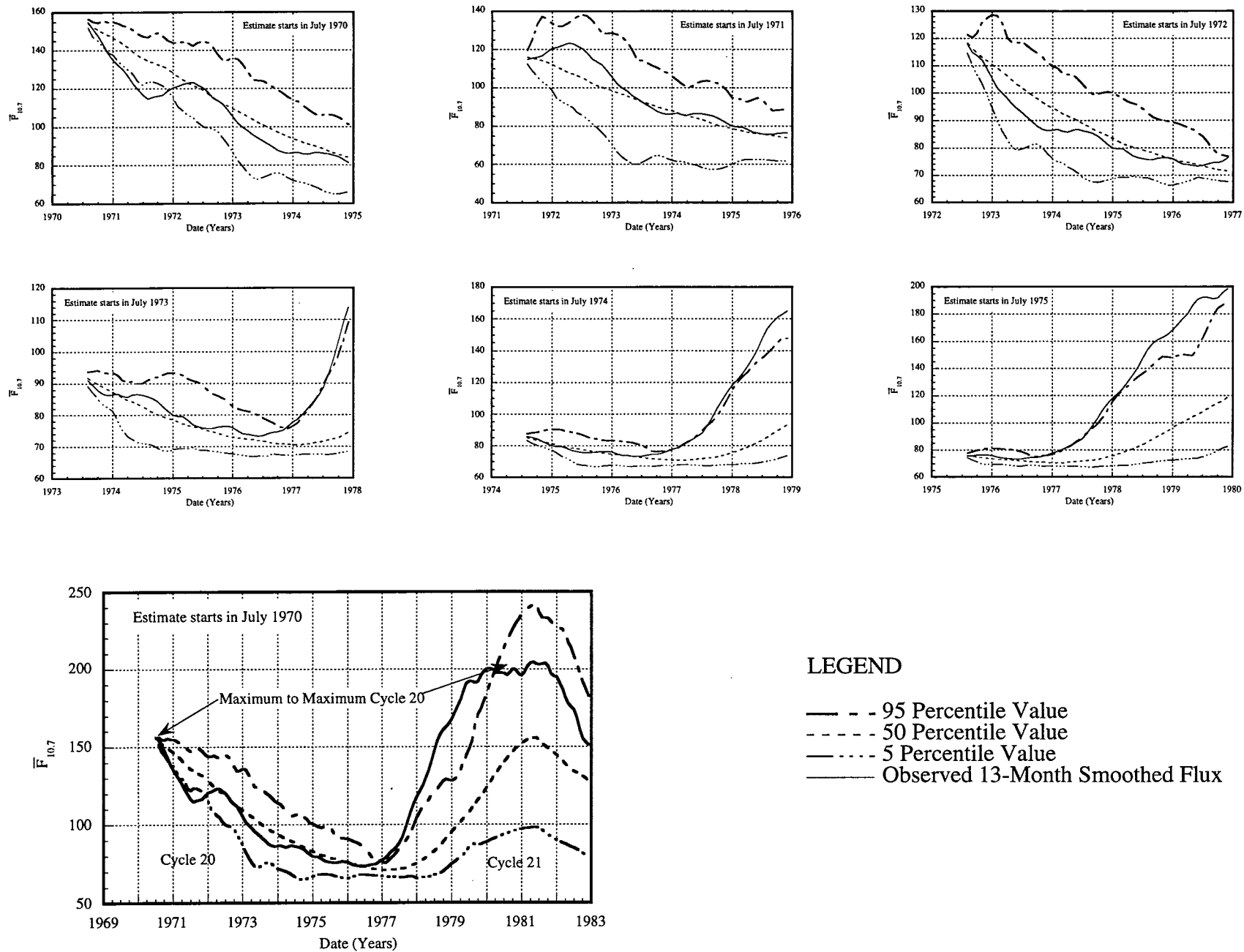


Figure F-2. Maximum to Maximum Cycle 20 MLLRT Model Results Evaluation.

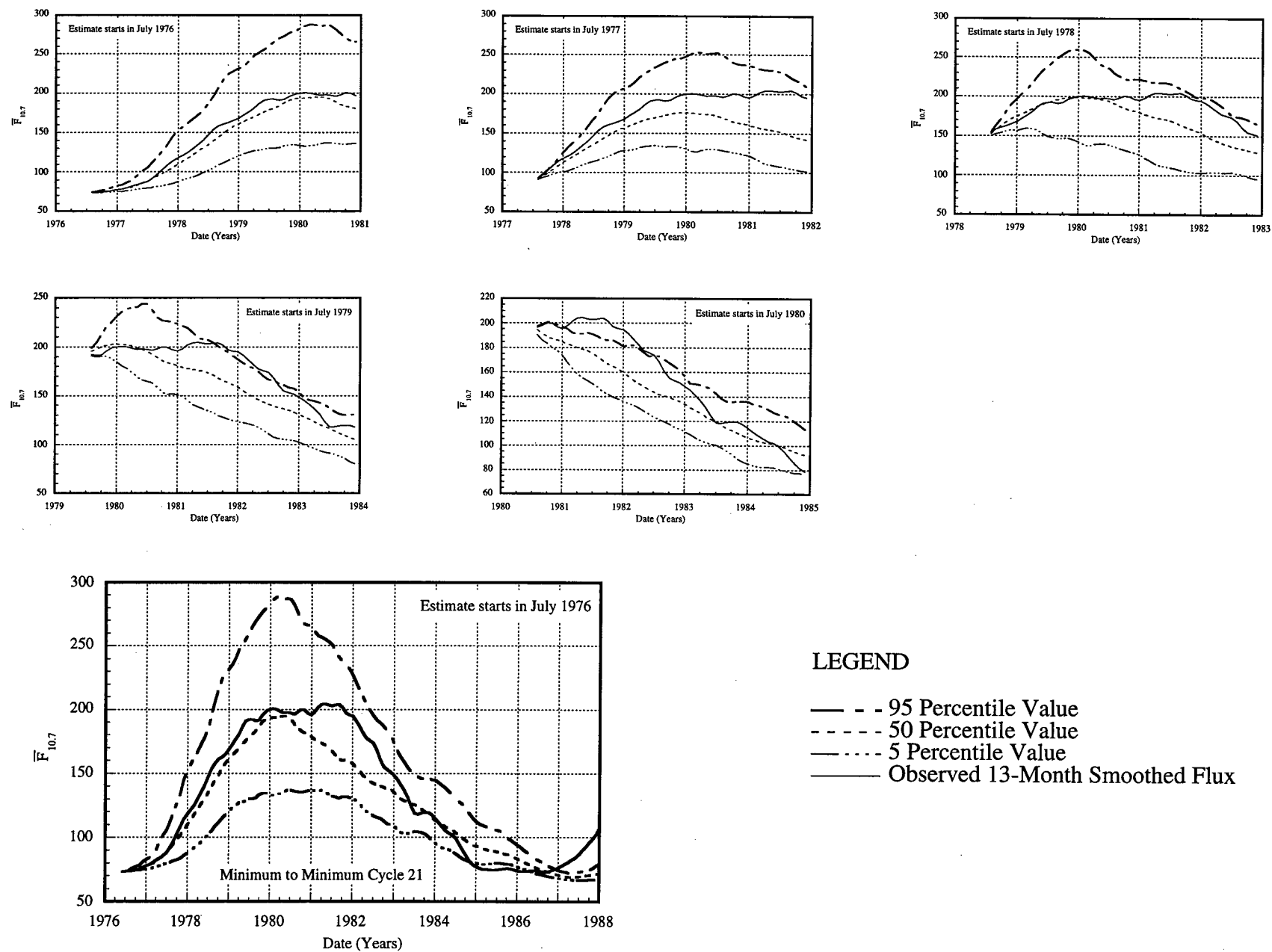


Figure F-3. Minimum to Minimum Cycle 21 MLLRT Model Results Evaluation.

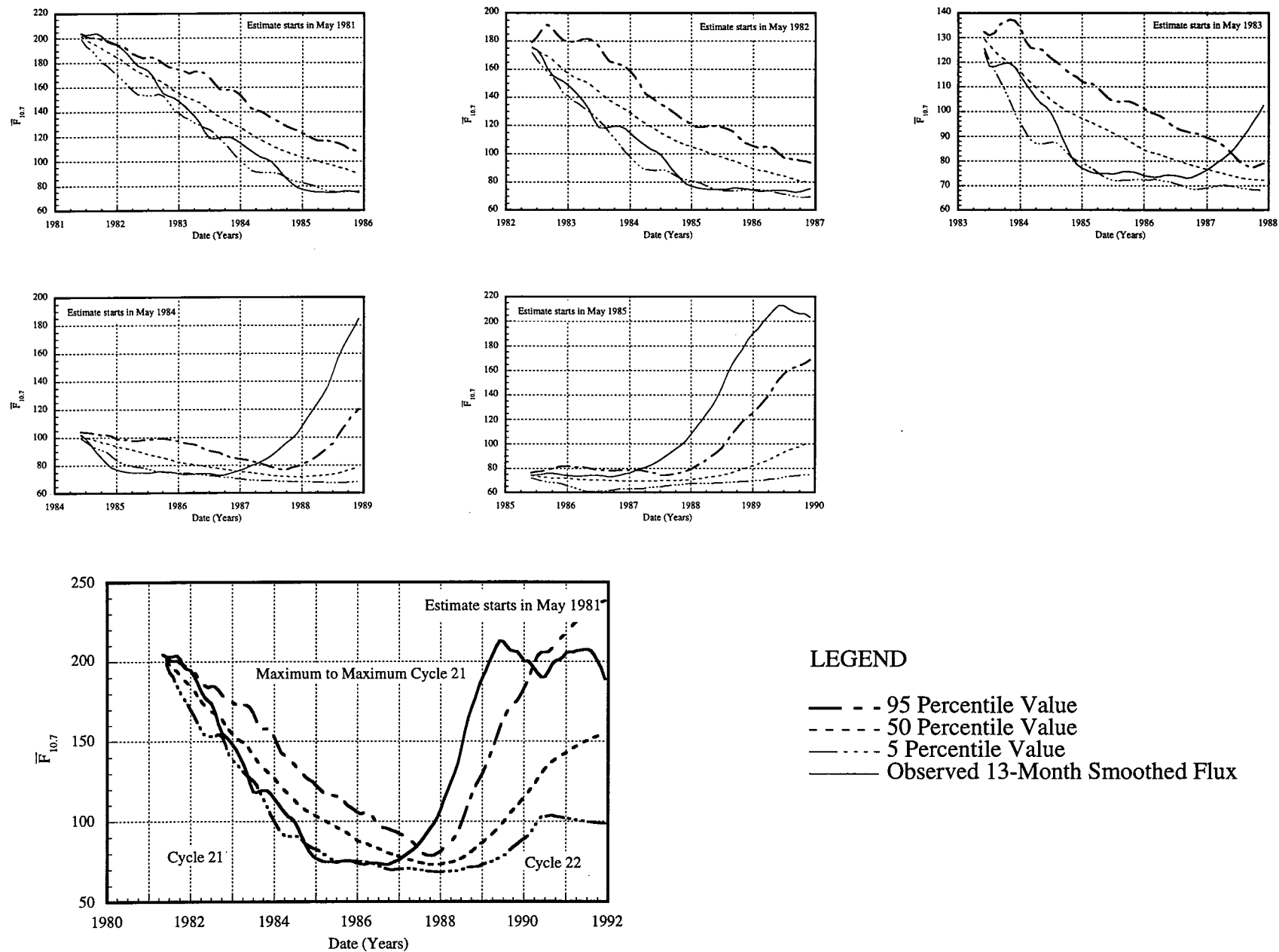


Figure F-4. Maximum to Maximum Cycle 21 MLLRT Model Results Evaluation.

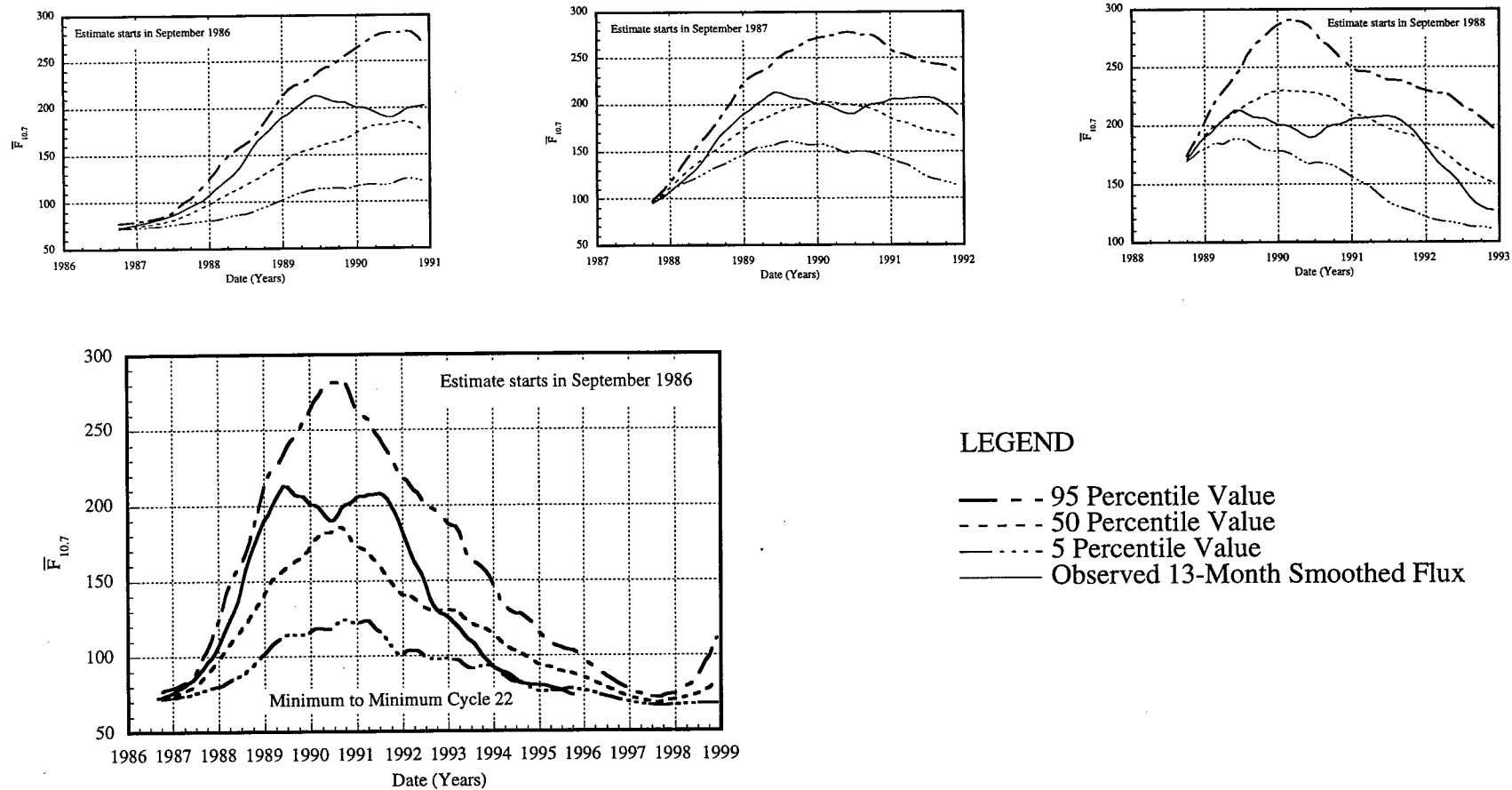
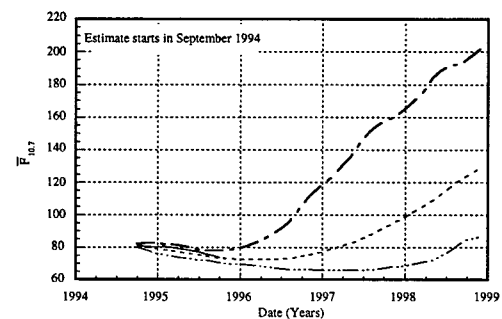
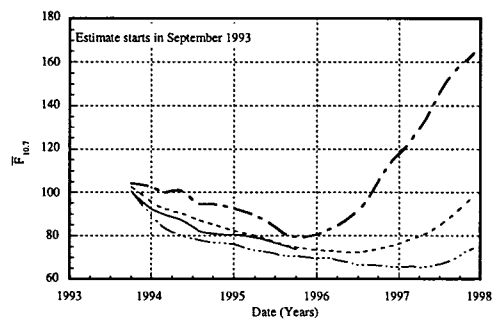
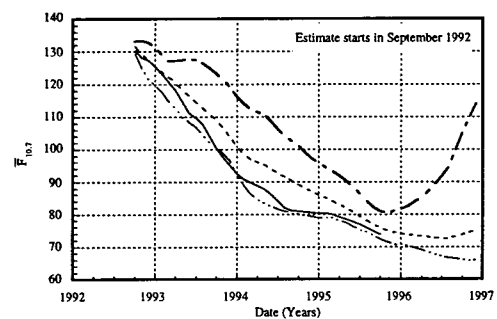
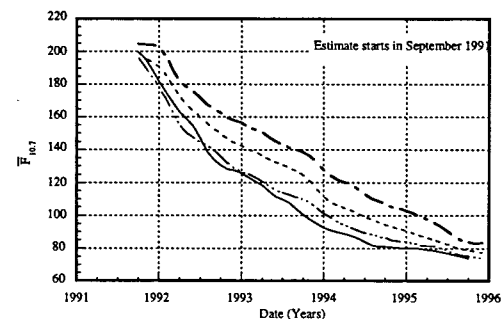
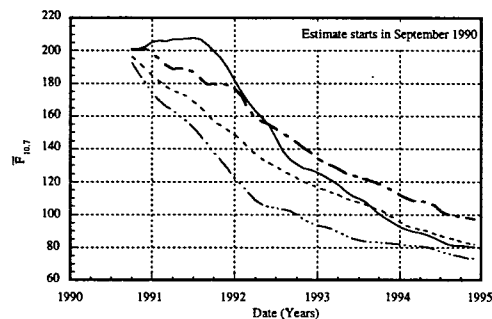
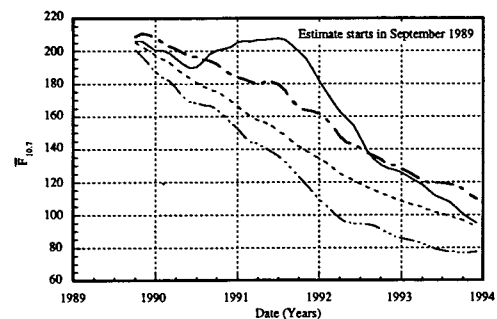
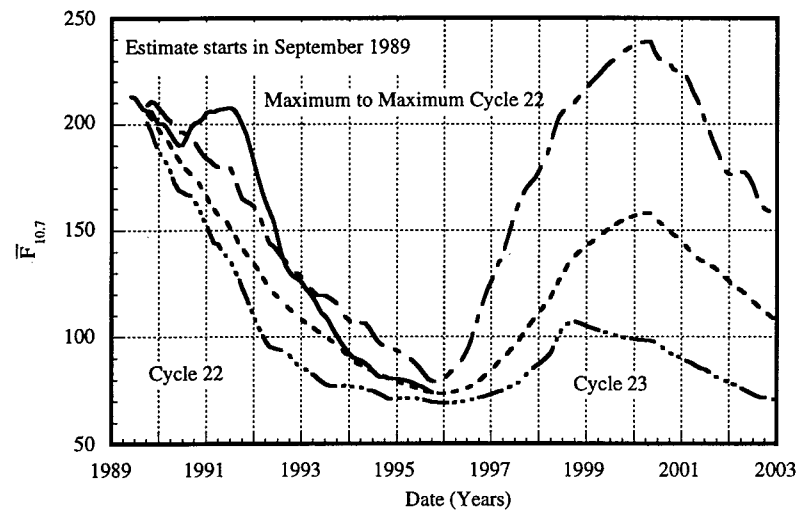


Figure F-5. Minimum to Minimum Cycle 22 MLLRT Model Results Evaluation.



F-7



LEGEND

- 95 Percentile Value
- 50 Percentile Value
- 5 Percentile Value
- Observed 13-Month Smoothed Flux

Figure F-6. Maximum to Maximum Cycle 22 MLLRT Model Results Evaluation.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operation and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503				
1. AGENCY USE ONLY (Leave Blank)		2. REPORT DATE September 1996		3. REPORT TYPE AND DATES COVERED Technical Memorandum
4. TITLE AND SUBTITLE Statistical Technique for Intermediate and Long-Range Estimation of 13-Month Smoothed Solar Flux and Geomagnetic Index			5. FUNDING NUMBERS	
6. AUTHORS K.O. Niehuss, H.C. Euler, Jr., and W.W. Vaughan*				
7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(ES) George C. Marshall Space Flight Center Marshall Space Flight Center, Alabama 35812			8. PERFORMING ORGANIZATION REPORT NUMBER M-815	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) National Aeronautics and Space Administration Washington, DC 20546			10. SPONSORING/MONITORING AGENCY REPORT NUMBER NASA TM-4759	
11. SUPPLEMENTARY NOTES Prepared by Systems Analysis and Integration Laboratory, Science and Engineering Directorate *University of Alabama in Huntsville, Huntsville, Alabama				
12a. DISTRIBUTION/AVAILABILITY STATEMENT Unclassified-Unlimited Subject Category 18			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) This report documents the Marshall Space Flight Center (MSFC) 13-month smoothed solar flux ($F_{10.7}$) and geomagnetic index (A_p) intermediate (months) and long-range (years) statistical estimation technique, referred to as the MSFC Lagrangian Linear Regression Technique (MLLRT). Estimates of future solar activity are needed as updated input to upper atmosphere density models used for satellite and spacecraft orbital lifetime predictions. An assessment of the MLLRT computer program's products is provided for 5-year periods from the date estimates were made. This was accomplished for a number of past solar cycles.				
14. SUBJECT TERMS natural space environment, solar activity, solar flux, geomagnetic index, statistics, estimates			15. NUMBER OF PAGES 81	
			16. PRICE CODE A05	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited	